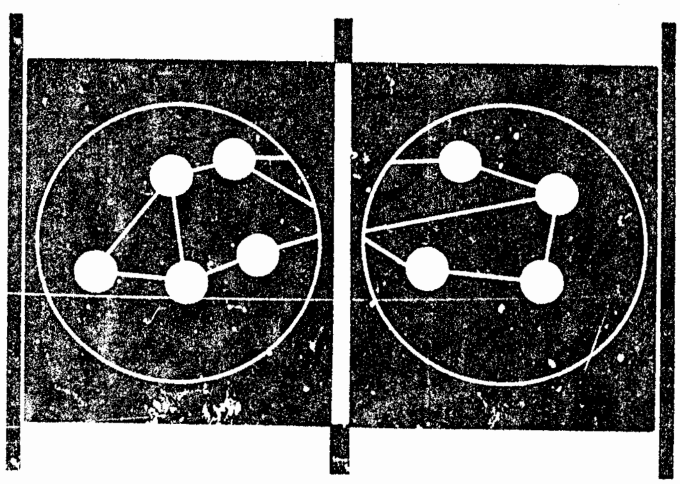
Local, Regional and Large Scale Integrated Networks

Sixth Semiannual Technical Report

VOLUME 2

RECENT ADVANCES IN GROUND PACKET RADIO SYSTEMS



NETWORK ANALYSIS CORPORATION

9

Sixth Semiannual Technical Report. no. 6,

February 1976

11

12 184p.

For the Project

6

Local, Regional and Large Scale Integrated Networks.

VOLUME 2.

RECENT ADVANCES IN GROUND PACKET RADIO SYSTEMS.

10

Principal Investigator: Howard Frank,

Co-principal Investigators: Israel Gitman
Richard Van Slyke

Contractor

NETWORK ANALYSIS CORPORATION
Beechwood, Old Tappan Road
Glen Cove, NY 11542
(516) 671-9580

ARPA Order No. 2286

15

Contract No. DAHC 15-73-C-0135

15

Effective Date: 13 October 1972
Expiration Date: 30 June 1977

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

Sponsored by

Advanced Research Projects Agency
Department of Defense

1473

389161

LB

SIXTH SEMIANNUAL TECHNICAL REPORT

TABLE OF CONTENTS

VOLUME 4

LOCAL AND REGIONAL COMMUNICATION NETWORKS - TECHNOLOGIES AND ARCHITECTURES

VOLUME 5

COMPUTATIONAL TECHNIQUES

SUMMARY OF VOLUME 2

TECHNICAL PROBLEM

Network Analysis Corporation's contract with the Advanced Research Projects Agency in the area of Ground Packet Radio Technology has the following objectives:

- To develop routing and initialization algorithms for packet radio networks under general deployment conditions.
- To develop flow and stability control algorithms and communication protocols.
- To evaluate and study performance tradeoffs of hardware and software alternatives in the development of the packet radio technology.
- To develop programs and study the reliability of packet radio networks.

GENERAL METHODOLOGY

The approach to the solution of these problems has been the simultaneous,

- Development of algorithms suitable for broadcast radio networks.
- Development of analytical models for evaluation of proposed algorithms and protocols.
- Programming of the most promising algorithms and design alternatives and performance of final evaluation using a large scale simulation of the packet radio network.

TECHNICAL RESULTS

The following accomplishments are presented in this report:

- A hierarchical organization of communication protocols in large networks has been proposed, and the functions performed within each protocol level have been identified. This protocol structure enables simple modifications of headers when a packet traverses several nonhomogeneous networks. The protocols of the packet radio network have been classified into the structure proposed.

- Hardware and software alternatives for the packet radio technology have been compared by simulation on the basis of relative global performance. The major conclusions follow. Significant improvements in performance are obtained when the maximum number of packet transmission per hop is variable, rather than constant. Devices with improved time capture receivers result in a significant increase in network throughput as well as improved stability characteristics. A system which uses a header checksum in addition to the packet checksum performs much better than one which uses a packet checksum only. For a 100 kb/s low data rate channel, a good choice for the high data rate channel is between 400 kb/s and 800 kb/s; no significant gain is obtained beyond 800 kb/s.

- Analytical models for several network configurations have been developed. The models were used to determine network stability and evaluate the significance of various stability control procedures. Several levels of stability control procedures, based on the analysis, have been proposed.

- Efficient simulation programs for reliability analysis of packet radio networks have been developed. The reliability of packet radio networks was evaluated for three different routing algorithms. Other factors evaluated are: the transmission power of devices, the number of packet radio stations, and the location of stations.

DEPARTMENT OF DEFENSE IMPLICATIONS

The results of this report and previous NAC accomplishments in the area of ground packet radio technology have long and short term implications for the Department of Defense. In the long term, the technology provides new capabilities related to fast communication network deployment and mobile networks as well as being an alternative cost-effective solution for applications presently implemented with other technologies. In the short term, the studies identified potential problem areas for which capital investment will be beneficial in terms of improving the technology and minimizing development time.

IMPLICATIONS FOR FURTHER RESEARCH

The major areas identified for further research include: development of routing and initialization algorithms for mobile and for distributed packet radio networks, evaluating error correction schemes, and development of programs for packet radio network design.

TABLE OF CONTENTS

TABLE OF CONTENTS

FIGURES

TABLE OF CONTENTS

FIGURES (cont'd)

TABLE OF CONTENTS

FIGURES (cont'd)

TABLE OF CONTENTS

FIGURES (cont'd)

TABLE OF CONTENTS

FIGURE (cont'd)

TABLE OF CONTENTS

FIGURES (cont'd)

TABLE OF CONTENTS

FIGURES (cont'd)

TABLE OF CONTENTS
TABLES

TABLE OF CONTENTS

TABLES (cont'd)

VOLUME 2

Chapter 1

PACKET RADIO COMMUNICATION PROTOCOLS

1. PACKET RADIO COMMUNICATION PROTOCOLS

1.1 HIERARCHICAL DATA COMMUNICATIONS PROCEDURES

An abstract "data communication system" consists of "entities" which "transmit" and "receive" "messages" using an abstract "channel" (Figure 1.1). Three types of procedures are required in a data communication system:

1. "Channel operation procedures",

2. "Channel validation procedures", and

3. "Channel initialization and maintenance procedures".

The channel operation procedures allow entities to transmit and receive messages to and from other entities via the channel (open loop). Channel validation procedures maximize the reliability of communication on the channel by doing some or all of the following:

1. Guarantee that at least one copy of a message sent from one entity to another in fact arrived.

2. Throw away duplicate copies of messages.

3. Guarantee that a "message" that is received is the same as the one sent.

4. Regulate the flow of messages so that neither the receiving entities, nor the channel are overloaded.

5. Maintain sequencing of messages, if necessary.

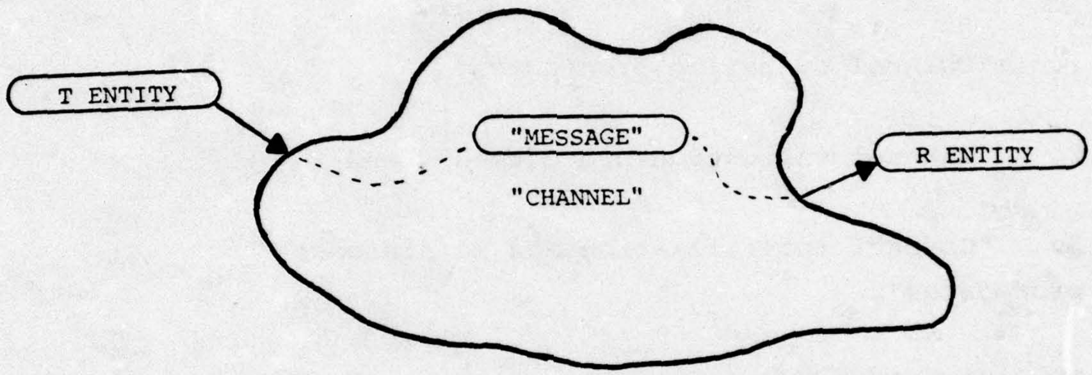


FIGURE 1.1: ABSTRACT DATA COMMUNICATION SYSTEM

Finally channel initilization, monitoring, measurement and maintenance procedures provide for the integrity of the channel.

Progressively, as modern data communications systems get larger they get more complex. A profitable way of analyzing and designing complex communication systems is hierarchically. (Data communication systems can even be considered recursively [POUZIN,1974]). When this is done, communication at one level in the hierarchy is carried out using lower level communication systems and in turn is used by higher level systems (Figure 1.2). To make all this more concrete, we turn to one possible implementation of protocols in a Packet Radio System in the next section. For other views of Packet Radio Networks see [KAHN, 1975], [FRANK, 1975].

The Packet Radio System described here differs in some ways both in objectives and details of implementation from the actual Packet Radio System being implemented by ARPA. The hierarchy we define is topological; that is, in the network of channels and switches which make up the system. Other partitions of communication procedures are possible and should not be confused with the one proposed here. For example, the partition could be according to levels of software in implementing computational devices, switching computers and the like; that is if a procedure "calls" another procedure the first is higher than the second in the hierarchy. A hierarchy defined along these lines is usually a refinement of a hierarchy defined in terms of communication function.

In a hierarchally structured communication system not all the validation functions need be performed at every level. Error checking, acknowledgments, sequencing and flow control are validation functions which can be performed at every level of a communications

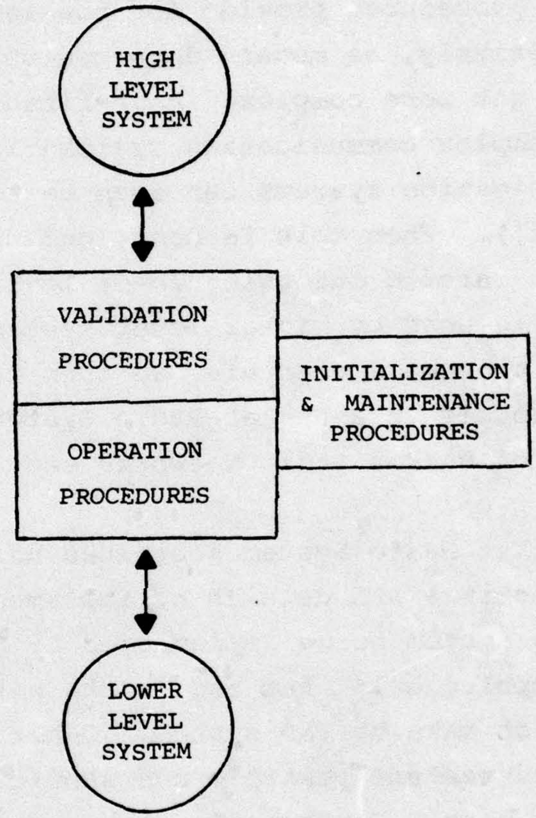


FIGURE 1.2: HIERARCHICAL DATA COMMUNICATION PROCEDURES

hierarchy. For acknowledgments, an end-tc-end acknowledgment at the highest level gives the utmost confidence. However, in unreliable systems delay can be radically reduced by acknowledgment verification at lower levels, thereby discovering errors more quickly after they occur.

This tradeoff between the efficiency of end-to-end acknowledgments at the highest level and the more responsiveness gained by lower level validation procedures is nothing but a new form of the throughput-delay tradeoff typical of virtually all communication systems. In the Packet Radio context, this tradeoff has been analyzed in [GITMAN, 1976].

1.2 THE PACKET RADIO SYSTEM

We consider a Packet Radio System (broadcast data network) designed to provide local collection and distribution of data over large geographical areas. The system should be economical, reliable, secure, and conservative of spectrum.

This Packet Radio Network (PRNET) includes three logical devices:

- Packet Radio Terminals,
- Packet Radio Stations, and
- Packet Radio Repeaters. (See Figure 1.3)

The Packet Radio Terminal consists of: (1) a device which sends or receives digital data; this includes TTY-like devices, CRT Terminals, sensors generating digital information, personal digital terminals, digital voice terminals, display terminals and computers - micro, mini, and maxi together with (2) an interface to the rest of the Packet Radio Network.

The Packet Radio Station performs accounting, buffering, directory, and routing functions for the overall system. In some applications the Packet Radio Station will also be the interface component ("Gateway") between the broadcast system and an external point-to-point network (Figure 1.3). As such it will have broadcast channels into the Packet Radio System and link channels into the point-to-point network.

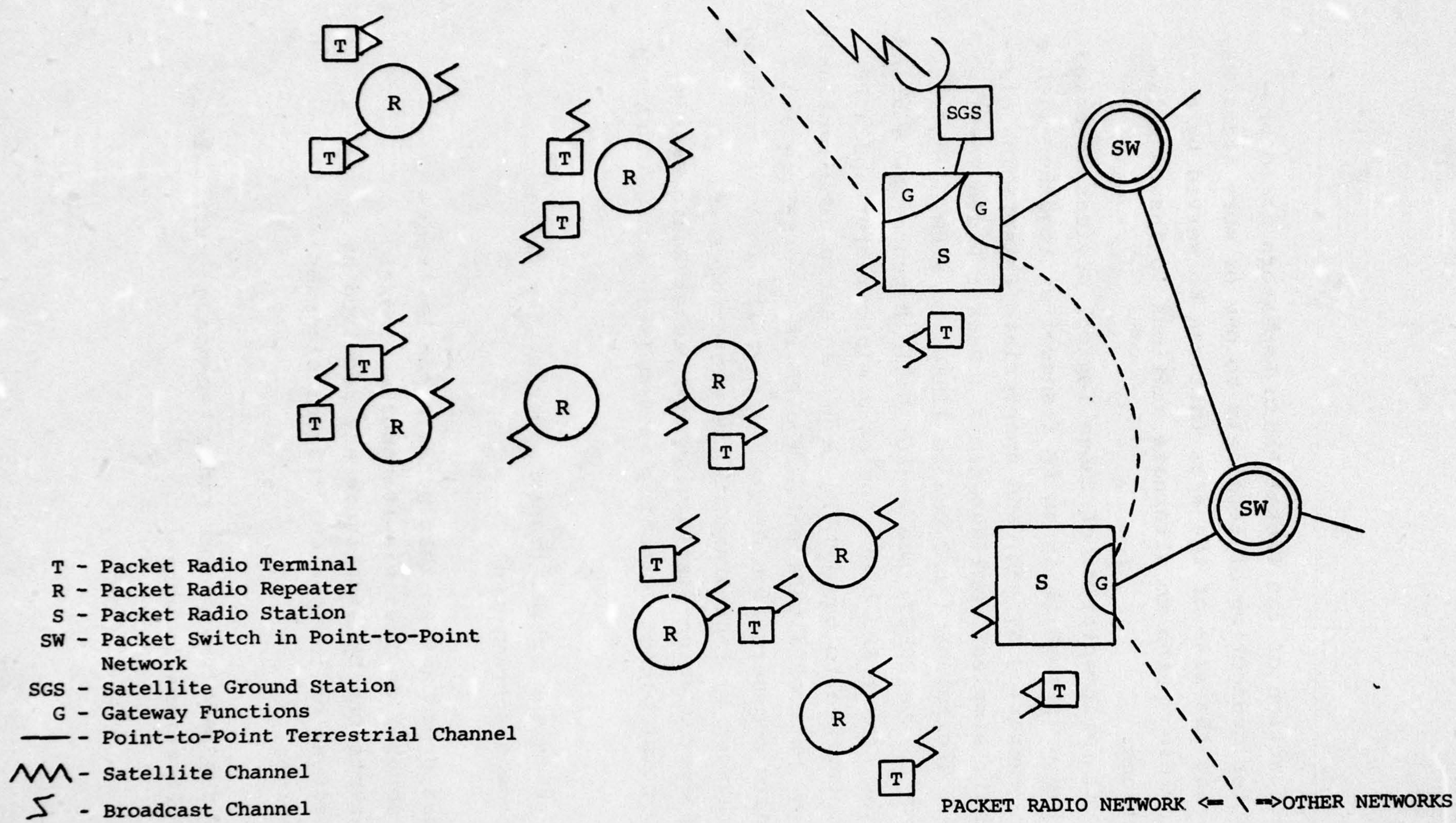


FIGURE 1.3: PACKET RADIO NETWORK

The basic function of the Packet Radio Repeaters is to provide a network for connection of terminals to one or more stations, thereby increasing the size of the area that can be served by a station and providing paths to alternate stations to insure reliable communications.

The physical implementation of these logical devices is based on the Packet Radio (PR) which is an RF transceiver together with a microprocessor, operating system, and appropriate interfacing electronics. In their simplest manifestations a Packet Radio Terminal is realized as a digital data I/O Device linked to a PR which acts as the interface to the PRNET; the Packet Radio Repeater as a stand-alone PR; and the Packet Radio Station as a minicomputer and a PR acting as the interface to the PRNET. A PR can serve simultaneously as a repeater and part of a terminal. The PR is a versatile and general radio data communication device in that its logical function can be easily changed by reprogramming the microprocessor.

The main features which distinguish the Packet Radio System from a point-to-point packet switching system (such as the ARPANET) are:

1. Devices in the system transmit packets by using a random access scheme, and

2. Devices broadcast so that packets can be transmitted to several devices simultaneously, and/or several packets can be simultaneously received by a receiver because of independent transmissions of several devices.

These features make the packet radio technology particularly suitable for applications in which:

1. Terminals are mobile, so that a broadcasting mode is necessary.

2. Terminals are located in remote or hostile locations where hardwire connections are not feasible.

3. Terminals have a high ratio of peak bandwidth to average bandwidth requirements (because the random access method allows the dynamic allocation of channel capacity without centralized control).

4. Terminals require little communication bandwidth so that hardwire connections are uneconomical.

1.3 PACKET RADIO NETWORK COMMUNICATION PROTOCOLS

A broad classification of PRNET protocols is given in this section. It is useful to classify protocols into hierarchy levels and/or according to clearly defined functions, because it enables one to associate a particular protocol (or function) with a hardware and/or software module which uses the protocol. Protocol classification also enables the design of *non-overlapping* sections of the packet (or packet headers) and association of each section with a function or protocol. This is particularly useful when a packet traverses different media (networks) since by properly placing the various headers, it enables one to strip off a header which is no longer needed or add a new header when needed so as to minimize packet overhead. Finally, protocol classification enables one to assign hardware and/or software development responsibility to different groups, minimizing interfacing problems.

The protocols used by the PRNET communication devices can be classified into three hierarchy levels associated with the traversing of a packet over a single hop, over the PRNET, and between processes of the end devices. These three levels can be further subdivided into functions of operation, validation, and initialization and maintenance.

Level I:	Inter PR Protocols
Level II:	Terminal - Station Protocols
Level III:	Inter-Process Protocols

The inter PR (Packet Radio) protocols include all the procedures applied to a packet from the time it is read out of a PR buffer for transmission until the end of packet transmission and resetting of PR parameters, and the procedures used from the time a packet is detected (including the detection process) until the packet is written into a PR buffer and PR parameters are reset. This level also includes the protocols which govern communication between a terminal, station, or host and its PR. In terms of the model depicted in Figure 1.1, the transmitters and receivers are PR's and the channel is the ordinary broadcast channel used by the devices in the first case. In the second case, the transmitting and receiving devices are either a station, host, or terminal and a PR. The channel will ordinarily be a wire.

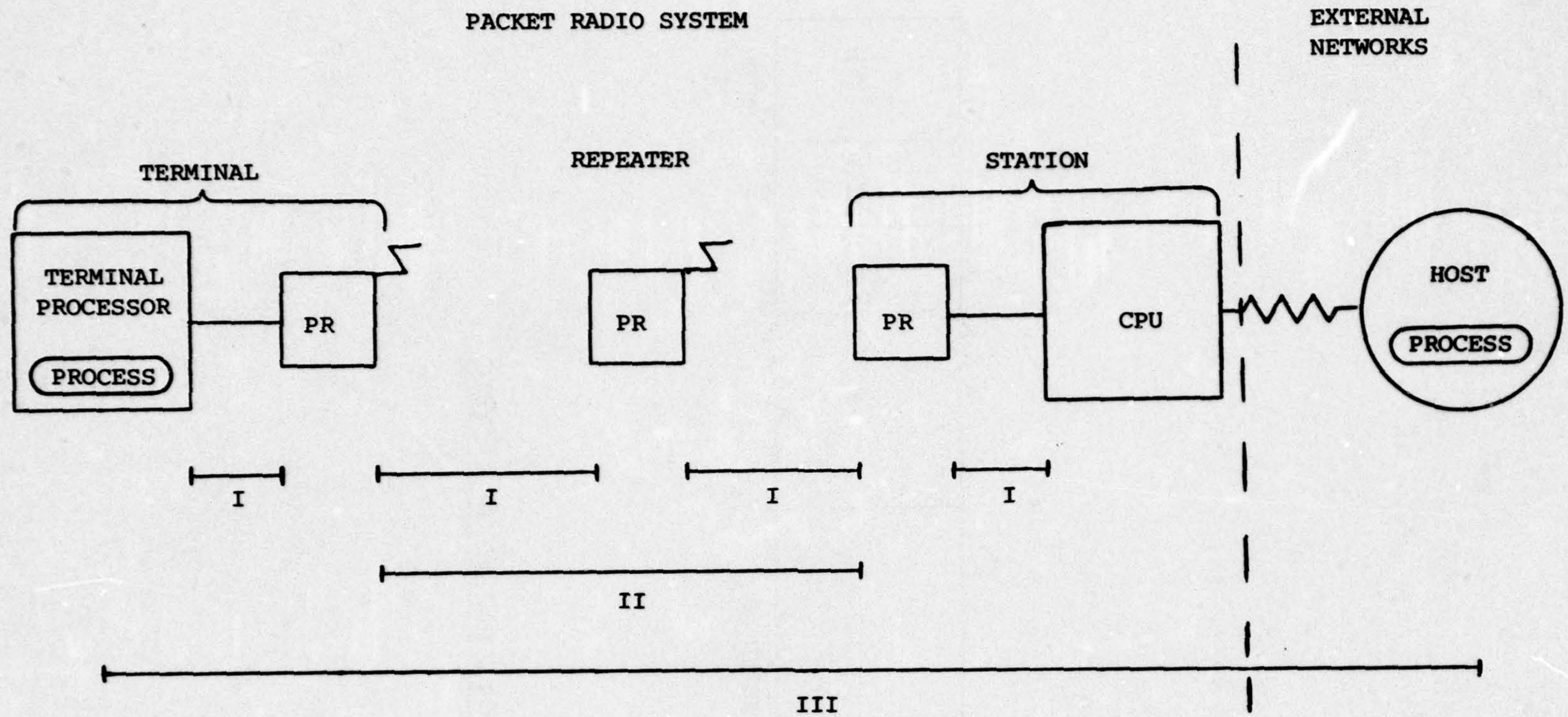
The terminal - station protocols are concerned with the efficient and reliable transport of packets between stations and terminals, stations and repeaters, and possibly stations to stations through the Packet Radio Network. Included are routing functions, flow control in the Packet Radio Network, verification of correct transmission, initialization and control, management and measurement of network operation and performance. In this case, the receivers and transmitters are terminals, hosts, or possibly the associated PR's depending on the implementation itself. The corresponding protocol in the ARPANET is where the transmitters and receivers are the origination and destination IMP's (or TIP's) in the communication, and the "channel" is the ARPANET itself. In neither case does this level of protocol take any cognizance of the mechanism by which the message traverses the hops or links within the network.

The inter-process protocols is the class used between the programs or processes of the origination device and programs or processes of the destination device; either of which or both may be in the packet radio network. These protocols establish and maintain the connection between the end processes, and include the set of rules for the exchange of messages between the particular end processes. An example of such a protocol is the INTERNET protocol of Cerf and Kahn [CERF, 1974], [CERF, 1975]. The inter-process protocols are transparent to the packet radio repeater and to the section of the packet radio terminal and station which interface with the PRNET. Figure 1.4 shows schematically the Protocol Levels for the PRNET.

Protocols are reflected in the packet as packet headers; the term "header" is used even though the information appears both at the beginning and the end of the packet. It is logically most consistent that the protocol information goes from the lowest to highest level from the outside of the packet to the inside. Figure 1.5 is a schematic of a proposed PRNET packet.

Ordinarily, in each of the three levels of hierarchy the "a" part of the header includes such information as the size of the packet, its type relative to the protocol level in question, and possibly its position in a sequence scheme. Ordinarily the "b" part contains a checksum for the appropriate level.

Let us briefly follow a packet from sending process to receiving process to see what happens to the header in transit. For concreteness, we will assume that the sending process is in a PR terminal and the receiving process is a host on the ARPANET. The initiating process passes the text along with the Level III header to the terminal. The terminal may subdivide the text into smaller packets and append the Level II header. The original text (or piece of it) and the Level III header will appear to the Level II protocol as undifferentiated "text". That is, the Level III header will not be recognized. Then the terminal



I - Inter PR
II - Terminal - Station
III - Interprocess

FIGURE 1.4: COMMUNICATION PROCEDURES HIERARCHY

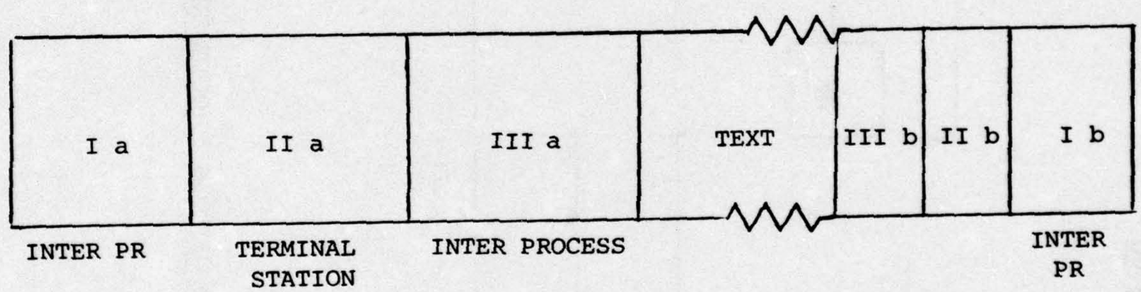


FIGURE 1.5: PACKET SCHEMATIC

attaches a Level I header corresponding to the protocol for the interface between terminal, host, or station and its associated PR. At the PR, the Level I checksum is verified if present, and the old Level I header is taken off the packet and a new Level I header which will, in general, have a different format corresponding to a Packet Radio hop - i.e., inter PR - is added. Eventually, the packet(s) reach a gateway to the ARPANET. At this point, the Level II header is replaced by the Level II header appropriate to the ARPANET (also, the format of the Level I header changes to the link protocol format of the ARPANET at this point). Finally, the host is reached, the message reassembled and the Level III checksum (if present) can be verified. Note that the Level III checksum offers end-to-end validation; however, in general (especially if the packets are divided) the sum cannot be verified until the very end of the transmission. If the PRNET turned out to be very unreliable and the delay in the ARPANET very long (say because of satellite channel delays), this could lead to very long expected delay time. In this situation, one would certainly also want Level I or Level II checksums.

If the headers corresponding to the different levels are configured as in Figure 1.5, then each level checksum is not disturbed by changes in lower level header fields.

Unfortunately, the idealized partition implied by Figure 1.5 is often very expensive in bits of header required. In practice, many fields are shared by the various levels. For example, if the headers of the various levels have a length known to each level protocol only, one packet length field is necessary. It can be shared by all levels.

VOLUME 2

Chapter 1

REFERENCES

REFERENCES

[CERF, 1974] Cerf, V. G., and R. E. Kahn, "A Protocol for Packet Network Intercommunication," IEEE Transactions on Communications, May, 1974, pp. 537-648.

[CERF, 1975] Cerf, V. G., "ARPA Internetwork Protocols Project Status Report," Technical Note #68, November, 1975, available from Defense Documentation Center, Arlington, Virginia.

[FRANK, 1975] Frank, H., I. Gitman, and R. M. Van Slyke, "Packet Radio System - Network Considerations," Proceedings of the National Computer Conference, May, 1975, pp. 217-231.

[GITMAN, 1976] Gitman, I., "Comparison of Hop-by-Hop and End-to-End Acknowledgment Schemes in Computer Communication Networks," to be published in the the IEEE Trans. on Communications.

[KAHN, 1975] Kahn, R. E., "The Organization of Computer Resources into a Packet Radio Network," Proceedings of the National Computer Conference, May, 1975, pp. 177-186.

[POUZIN, 1974] Pouzin, L., "An Integrated Approach to Network Protocols," INWG Protocol Note #12, November, 1974, available from Stanford Electronics Laboratories, Stanford University, Stanford, California 94305.

VOLUME 2

Chapter 2

PERFORMANCE EVALUATION OF PACKET RADIO NETWORKS

2. PERFORMANCE EVALUATION OF PACKET RADIO NETWORKS

2.1 INTRODUCTION

The performance of the packet radio network can be improved by software means (e.g., improved routing and communication protocols) or by hardware means (e.g., increasing the channel capacity in the repeater-station network or improving the capture of receivers). A quantitative demonstration of the increase in system capacity and/or the reduction in the delay for each specific improvement enables more objective decision making regarding the means one should choose. It is not the objective of this chapter to establish absolute values of performance for the various system modifications considered. Rather, the objective is to evaluate specific hardware and software design alternatives. The various modifications are compared on the basis of relative performance obtained by simulation. The basic elements which define the system simulated and the criteria used for comparison are presented in Section 2.2.

One of the characteristics of the packet radio network is that packets which cannot be delivered within an assigned amount of resources are discarded. This prevents lockups resulting from the unavailability of repeater buffers on the one hand, and, on the other hand, it reduces the possibility of local traffic instability when random access schemes are used. Furthermore, there seems to be no rationale in keeping the packet in the network for a long period of time since another copy of same will be introduced via the mechanism of end-to-end retransmission; the latter ends up competing with the former for network resources. The Maximum Number of Transmissions (MNT) per hop is a measure of the amount of resources assigned to a packet for traversing in the radio network. In Section 2.3, we compare network performance as a function of MNT and recommend MNT values for practical networks [NAC, 1975a].

In Section 2.4 we compare the performance of the packet radio system with Zero Capture receivers against a system with Perfect Capture receivers. The objective is to determine whether improved reception characteristics will significantly improve system performance and to identify other properties resulting from improved reception [NAC, 1975b].

An alternative way to improve reception is by using a checksum at the end of the packet header in addition to that used at the end of the packet. The advantage gained by using the header checksum is by being able to utilize the header information in cases in which the portion of the header is correctly received whereas the whole packet is received in error. The Hop-by-Hop Acknowledgment (HBH Ack) scheme in the packet radio system is based on the correct reception of a packet which is already stored in the receiving device, when the packet was transmitted by a device closer to the destination. Hence, in cases in which it is a packet with text, there is a higher probability for correctly receiving the header section than the probability of correctly receiving the entire packet. This is true in systems which use non-slotted ALOHA [ABRAMSON, 1970] or carrier-sense multiple access [KLEINROCK, 1975] schemes. The performance of packet radio systems with and without a checksum at the end of the header are compared in Section 2.5.

It has been demonstrated [FRANK, 1975] that a system with two data rate channels, a low data rate for communication between terminals and repeaters (or stations) and a high data rate for communication in the repeater-station network, performs significantly better than a system with a single data rate channel. One of the hardware design alternatives is to determine a proper value for the high data rate, for a given low data rate channel. In Section 6, we attempt to answer this question by studying the performance of the packet radio network as a function of the ratio of the high to low data rate channels.

2.2 DEFINITION OF THE SYSTEM AND PERFORMANCE MEASURES

A detailed description of the system simulated is given in [NAC, 1974]. The basic elements which define the system studied are stated below:

- Common channel two data rates. A low data rate for communication between a terminal and a repeater (or station), and a high data rate in the repeater - station network. A packet on the high data rate channel can interfere with a packet on the low data rate channel, and vice versa.

- The topology consisted of 1 station and 48 repeaters. The connectivity was moderate, which resulted in a network of many hops (up to 7). The location of repeaters and stations and the radio connectivity are shown in Figure 2.1. Terminals are introduced at random times and are placed in random locations in the plane "covered" by repeaters and stations. A terminal performs a short interaction of sending and receiving a few packets and then departs from the system. The rate at which terminals are introduced and the amount of communication depend on the traffic offered to the system and are controlled by parameters.

- The routing was hierarchical with restricted alternate routing; the alternate routing enables only forward transmission of a packet bypassing only one failed or busy repeater before returning to the established path [GITMAN, 1976]. The radio links assigned for routing form a tree structure. Figure 2.2 shows the structure of the network studied.

- The channel access scheme is non-persistent, non-slotted, carrier sense [KLEINROCK, 1975].

- Other parameters used in these experiments are a maximum of 3 end-to-end transmissions and a timeout of 60 slots (packet transmission time on low data rate) between end-to-end transmissions.

The following are the performance measures used for system evaluation and comparison:

- System throughput/system input rate. The throughput includes the acknowleged information packets. Hence, the above ratio is a lower bound on network performance.

- Percentage of total loss. The loss includes packets of terminals which are blocked and packets which are not delivered to the destination after the maximum number of end-to-end transmissions.

- Delays. The round trip delay averaged over all packets, independent of the number of hops from the station was used.

- Average buffer occupancy. The average number of occupied buffers in the entire repeater network is another measure used to compare the systems. More information about buffer occupancy is shown in the form of the number of packets as a function of time.

The time scale used in this report is the transmission time of an information packet on the low data rate channel, or a slot. If it is assumed that the size of an information packet is 2,000 bits and the low data rate channel is 100 Kbps, a slot time would be 20 msec. The size of short packets (ETE ack, search, response to search) used in the simulation is 10% of an information packet. Similarly, the traffic rates are in percentages of the low data rate channel; thus, a throughput of 30% means a throughput of 30 Kbps of bits in information packets.

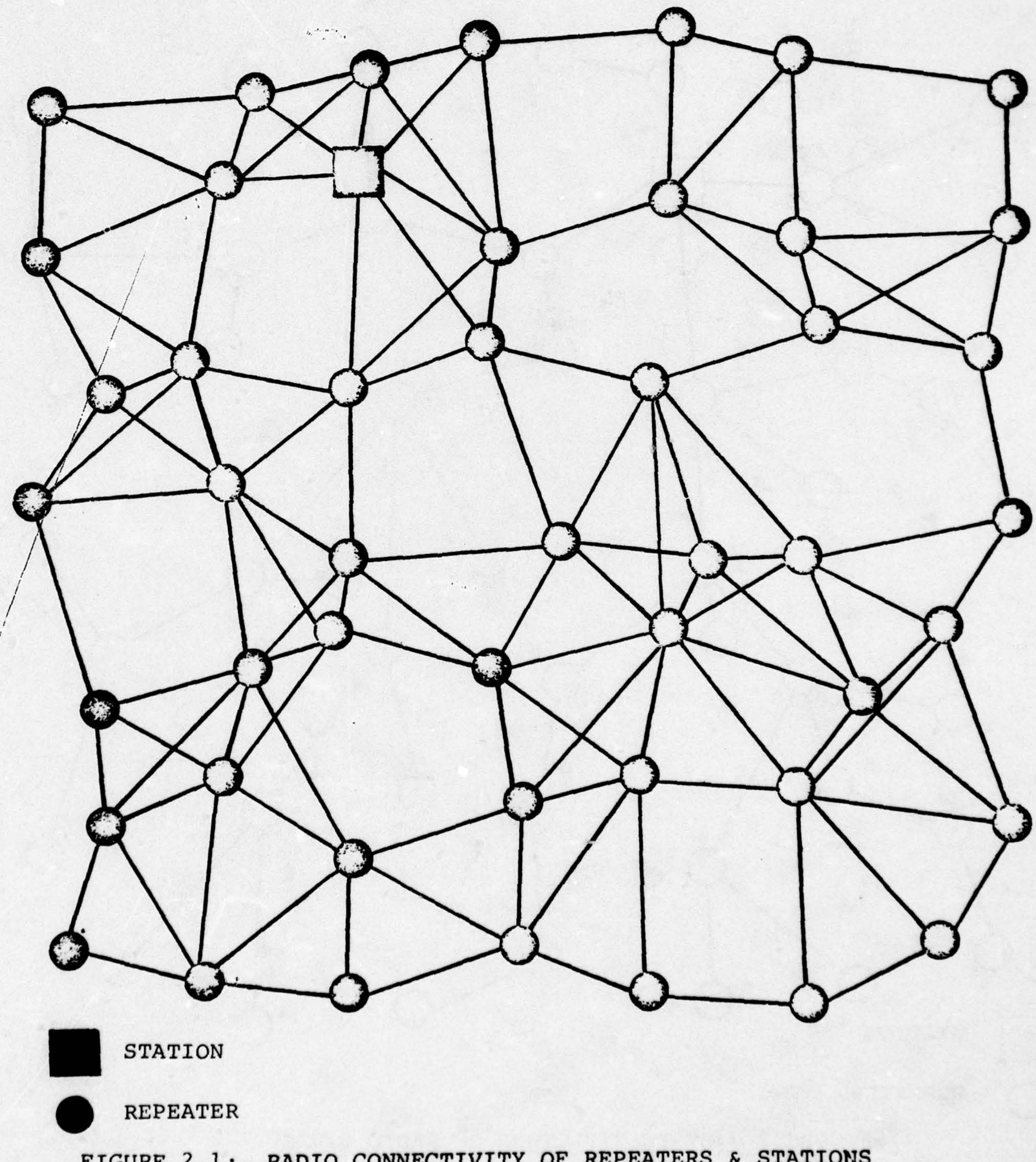


FIGURE 2.1: RADIO CONNECTIVITY OF REPEATERS & STATIONS

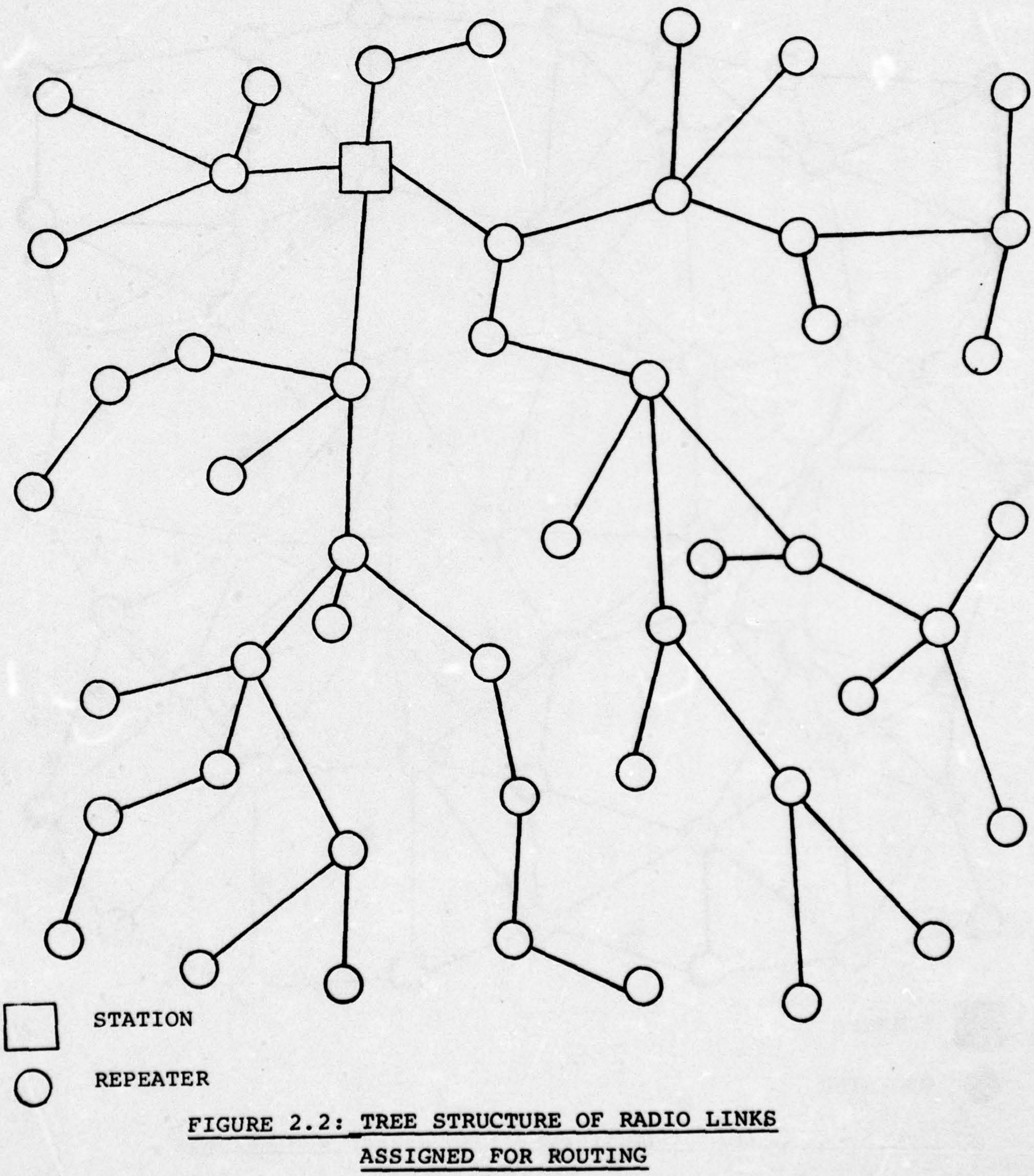


FIGURE 2.2: TREE STRUCTURE OF RADIO LINKS ASSIGNED FOR ROUTING

2.3 MAXIMUM NUMBER OF TRANSMISSIONS FOR REPEATERS AND STATIONS

In this section, we report on a study for evaluating network performance as a function of the amount of resources assigned to a packet. In the present implementation of the packet radio network, the only parameter which determines the resources allocated to a packet per end-to-end transmission is the Maximum Number of Transmissions (MNT) per hop that a device uses before discarding the packet. This is because the routing has the property of being shortest path, the alternate routing is restricted (thus one cannot limit the number of hops), and the packet does not carry a time stamp which can be limited.

The system used for the study is that defined in Section 2.2, with a 4 to 1 ratio of data rates, a zero capture receiver, and a header checksum.

In the study, all system parameters were kept constant and the performance was studied as a function of MNT. The values of MNT used were: 3, 6, 9, 15, ∞, and a variable number which was a function of the number of hops from the station. For the variable MNT the formula used was [10-(Hierarchy Level)]. Thus, the station used an MNT of 9; repeaters one hop away from the station used the value of 8, etc. The reason for this variable MNT assignment are theoretical results showing that the average number of transmissions before success increases with the traffic level, and previous simulation results and analysis which demonstrated that the traffic bottleneck is near the station and that the traffic level decreases with the distance (in hops) from the station. The value of infinity (∞) was used as a reference value. It also simulates the case in which packets are not discarded until successfully forwarded, as is done in point-to-point store-and-forward networks.

For each value of MNT, the system was run with two values of offered rate: 15% and 30%. The number of terminals simulated for the various tests ranged from 50 to 100. The average number of hops of the terminals from the station (measured) was 4 to 4.5.

Table 2.1 summarizes the performance of all the simulation runs. One can see that all the systems, apart from that with MNT = ∞, perform well when the offered rate is 15%. The unsatisfactory performance of the system with MNT = ∞ can also be seen in Figure 3, where the throughput and buffer occupancy are shown as a function of time. Although one can observe the same qualitative differences for the 15% offered rate as the ones for 30% offered rates, the absolute differences in performance seem to be too small to justify conclusions. Hence, the runs with 30% offered rate are used for comparison.

Figures 2.4, 2.5, and 2.6 compare the various systems in terms of performance measures defined in Section 2.2 as a function of MNT.

The comparison of the system for all values of MNT demonstrates that the variable MNT scheme shows much better performance in all measures used. There is a large difference in performance between the variable MNT and any of the fixed MNT cases. For example, the average round-trip delay for the variable MNT case is 17.04 slots whereas for other cases it is 34.12, 33.8, 39.0, 34.12, and 29.75 for MNT = 3, 6, 9, 15, and ∞, respectively. Figures 2.7 through 2.9 show the throughput and buffer occupancy as a function of time for the 30% offered rate. These figures immediately reveal the difference in performance.

The comparison of systems with MNT of 3 or 6 against systems with MNT of 9 or 15, shows that the former perform better. The only measure in which the system with MNT of 15 is better than

that with MNT of 3 is the average round trip packet delay per hop (8.12 vs. 8.66); however, this is obtained for a lower throughput (22.91% vs. 27.75%, respectively). In all other measures drawn in Figures 2.4 through 2.6, the performance of MNT of 3 or 6 is better than that of MNT 9 or 15.

In comparing MNT of 3 and 6, neither case is uniformly better than the other. The system with MNT of 3 shows better performance in the ratio of throughput to input rate, whereas that with MNT of 6 shows better performance in delay.

It is somewhat surprising that the system with MNT of 15 performs generally better than that with MNT of 9. However, the difference on performance is small; furthermore, the comparison is reversed in favor of the system with MNT of 9 for the runs with 15% offered rate.

As a final note, it is pointed out that the maximum throughput obtained in these experiments is higher than what was previously obtained with the same network topology and zero capture receivers. In previous experiments [FRANK, 1975], a throughput of 27.5% was obtained when the ratio of data rate was 5 to 1. The performance obtained in this series of experiments with the variable MNT is 29.12% throughput at the relatively low average round trip delay of 17.04 slots (approximately 340 milliseconds at 100 kilobits per second and 2,000 bit packets). Other performance measures such as the total loss and the buffer occupancy are also very good. This can be seen in Table 2.1 and Figure 2.9. This implies that the capacity of a single station packet radio network can be increased above 30% when protocols are improved and the delay requirements are relaxed.

	MAXIMUM NUMBER OF TRANSMISSIONS [MNT]	INPUT RATE [%]	THROUGH-PUT [%]	TOTAL LOSS [%]	ROUND TRIP DELAY AVERAGE [SLOTS]	ROUND TRIP DELAY AVERAGE/HOP [SLOTS]	AVERAGE BUFFER OCCUPANCY [PACKETS]
OFFERED RATE 15%	3	15.83	14.32	0	8.86	2.06	3.20
	6	13.59	13.41	0	8.81	2.17	2.57
	9	18.75	18.04	0	9.50	2.16	3.41
	15	16.83	16.18	10.89	11.13	2.65	3.80
	Variable MNT	15.17	14.75	0	8.18	1.98	2.40
	∞	12.24	6.34	27.87	10.83	2.55	6.59
OFFERED RATE 30%	3	32.83	27.75	6.34	34.12	8.66	7.34
	6	27.34	21.79	12.24	33.38	7.15	9.40
	9	30.37	20.85	16.49	39.00	9.22	11.65
	15	34.03	22.91	14.18	34.12	8.12	11.08
	Variable MNT	30.96	29.12	2.40	17.04	3.93	5.16
	∞	23.00	12.98	6.58	29.75	7.07	15.45

TABLE 2.1: SUMMARY OF PERFORMANCE MEASURES FOR THE MAXIMUM NUMBER OF TRANSMISSIONS STUDY

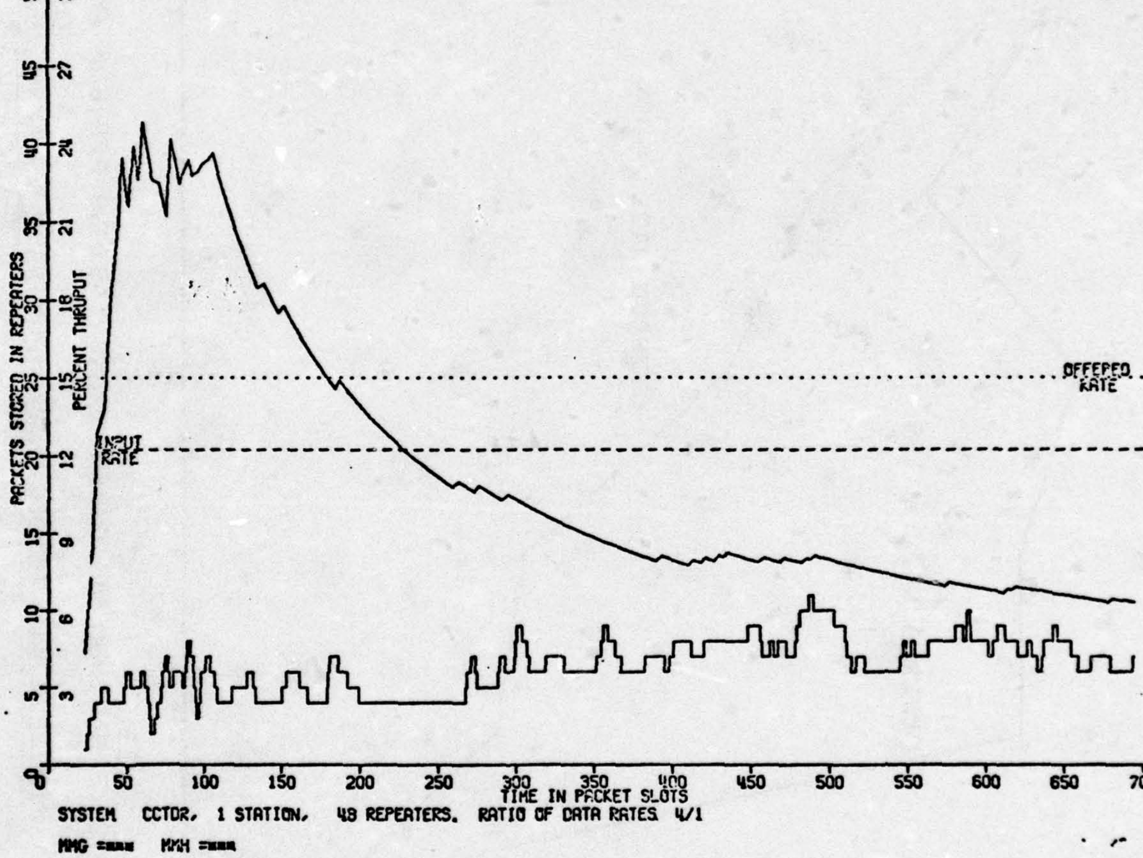


FIGURE 2.3: COMPUTER OUTPUT FOR MNT = ∞

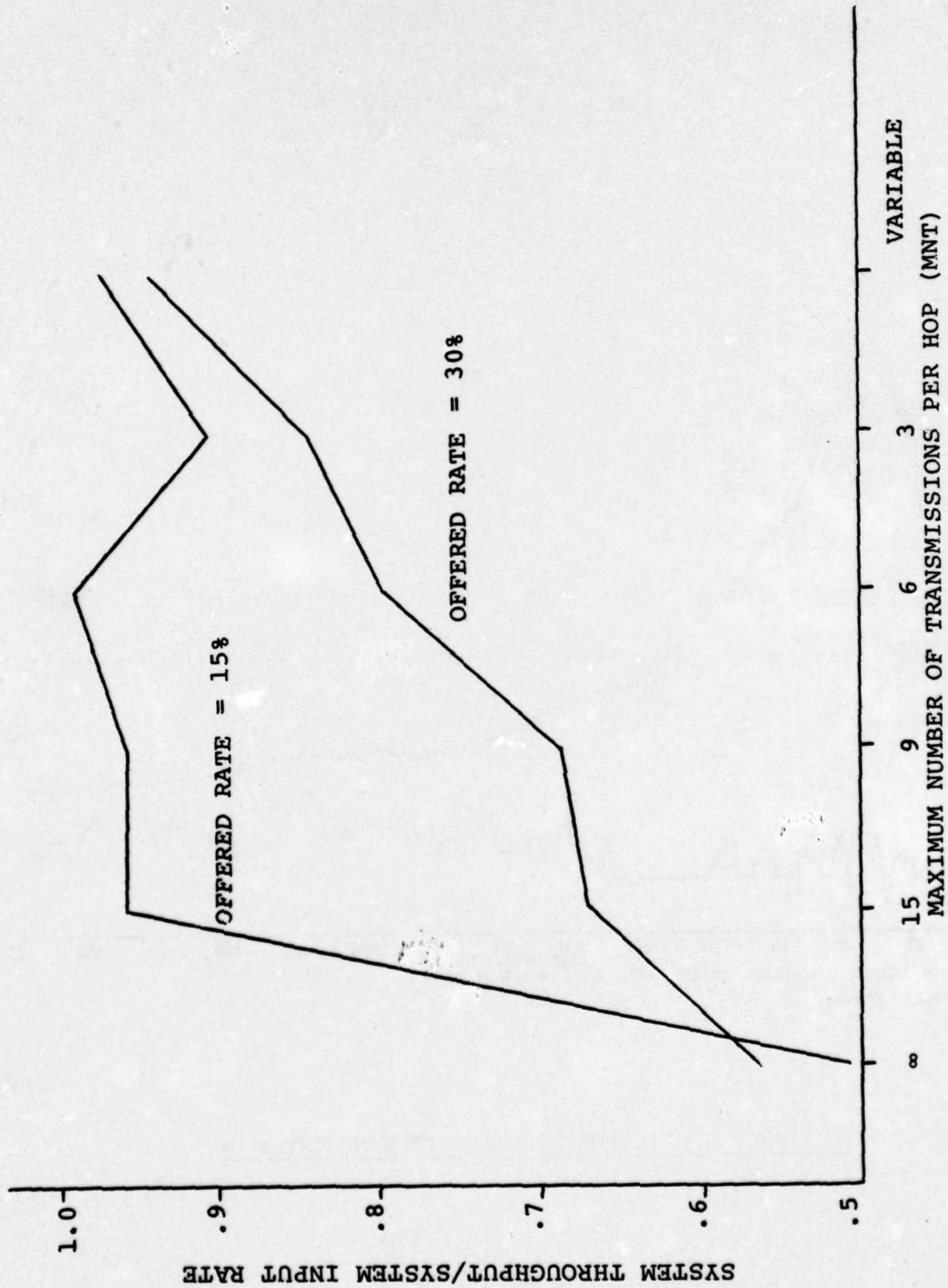


FIGURE 2.4: THROUGHPUT/INPUT RATE VS. MNT

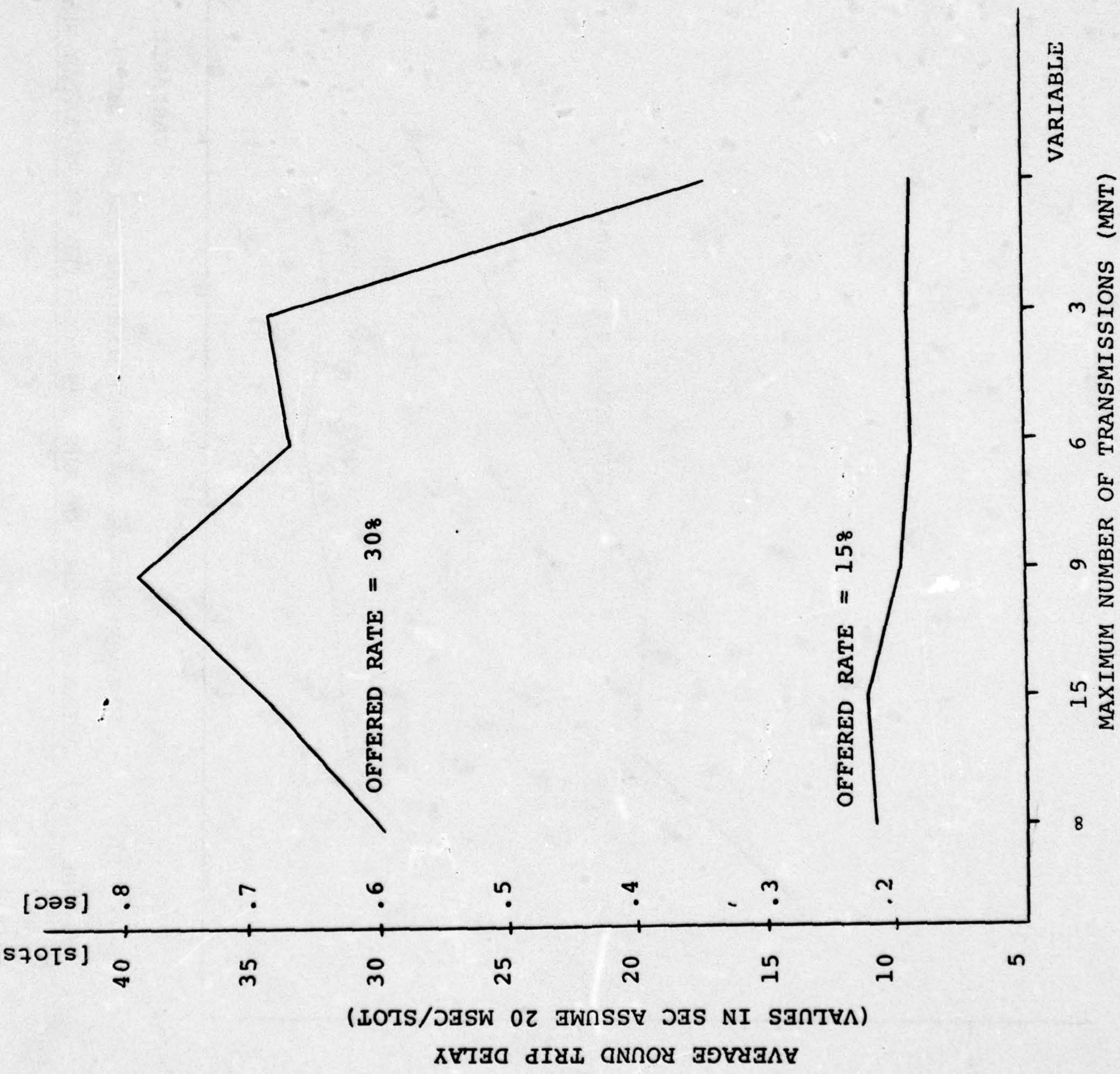


FIGURE 2.5: AVERAGE ROUND TRIP DELAY VS. MNT

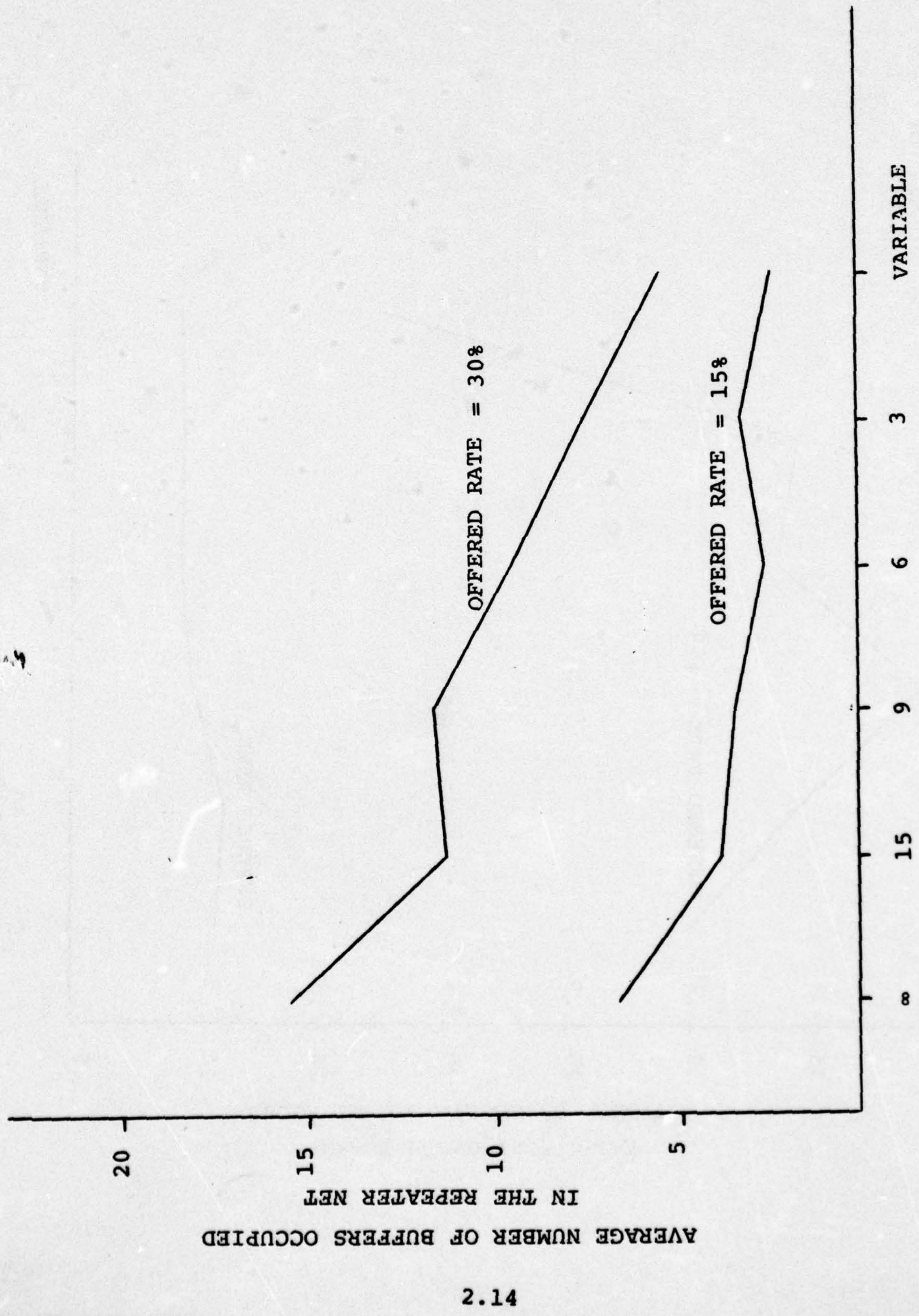
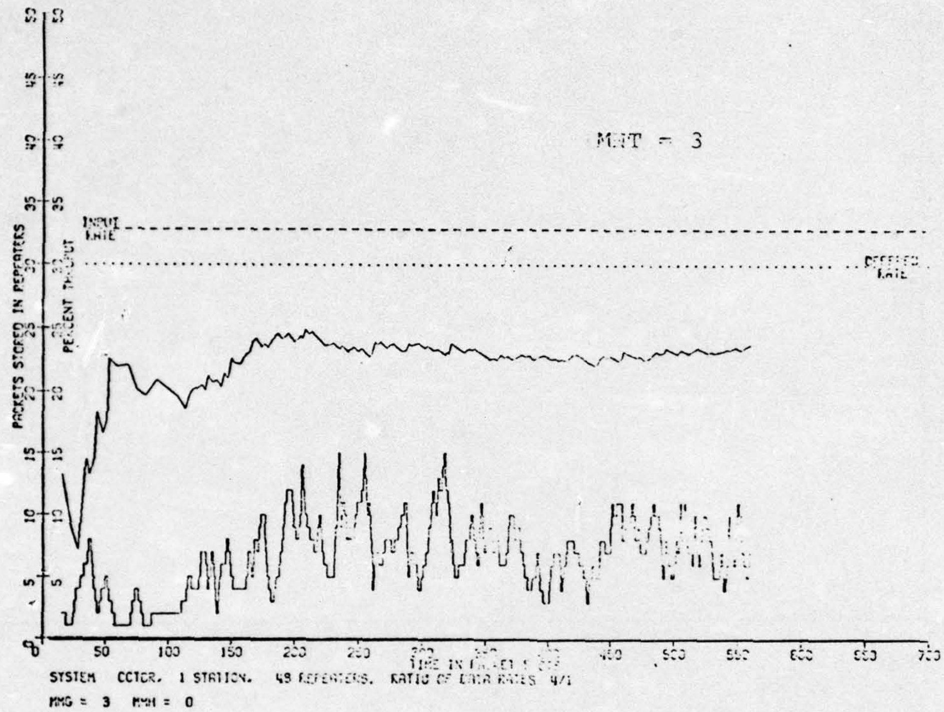


FIGURE 2.6: AVERAGE NUMBER OF BUFFERS OCCUPIED IN REPEATER NET VS. MNT



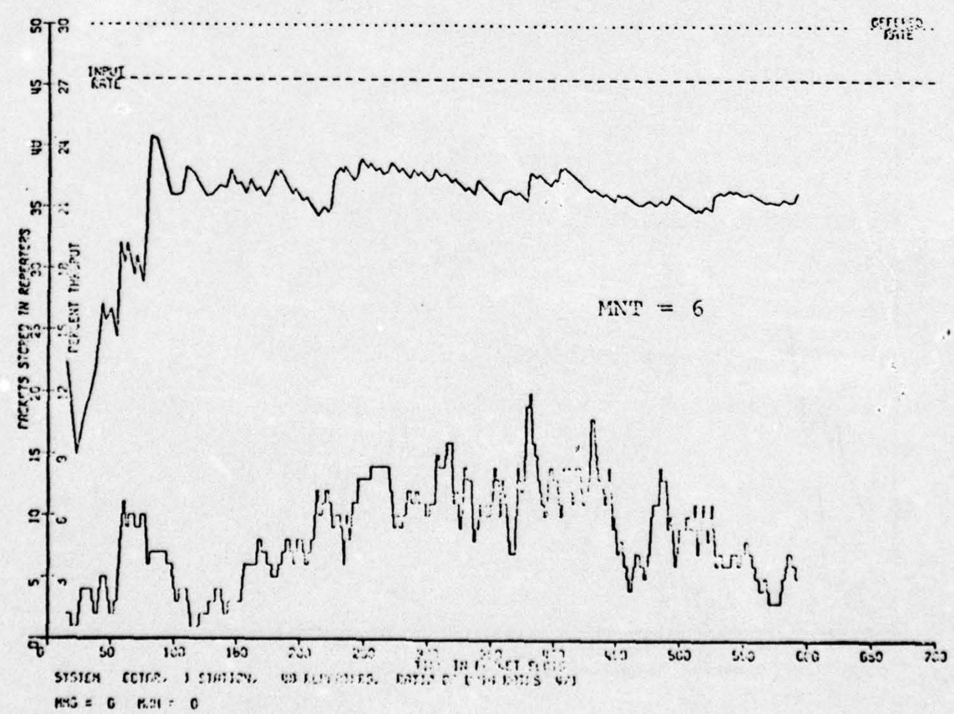
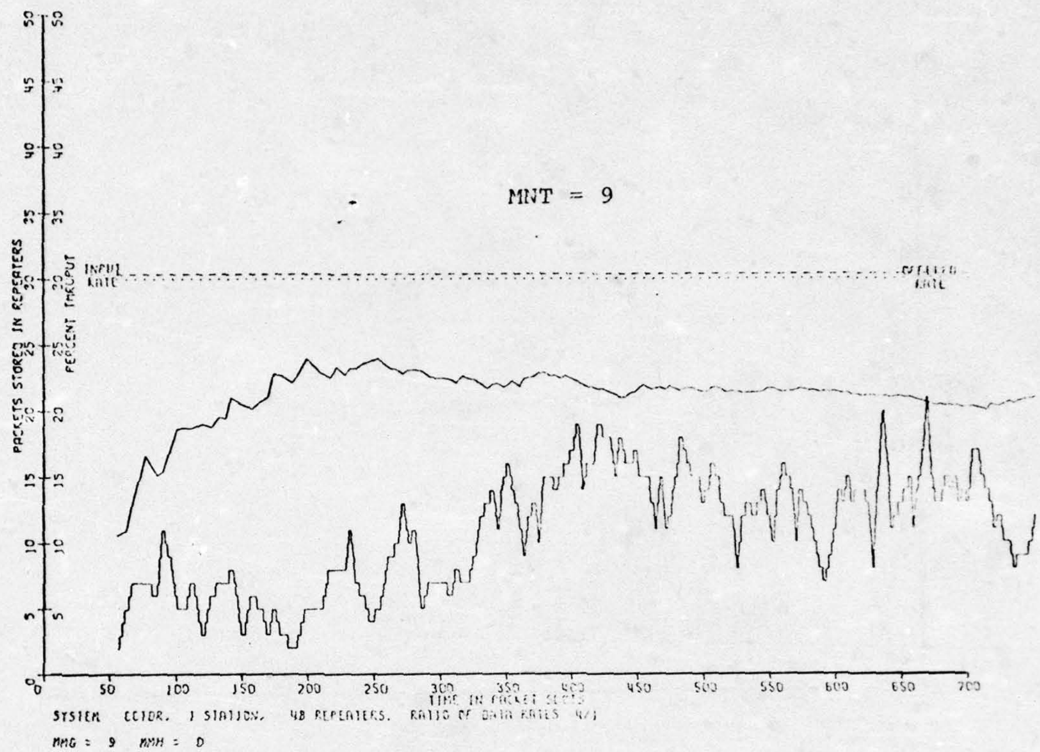


FIGURE 2.7: COMPUTER OUTPUT FOR MNT = 3 AND MNT = 6



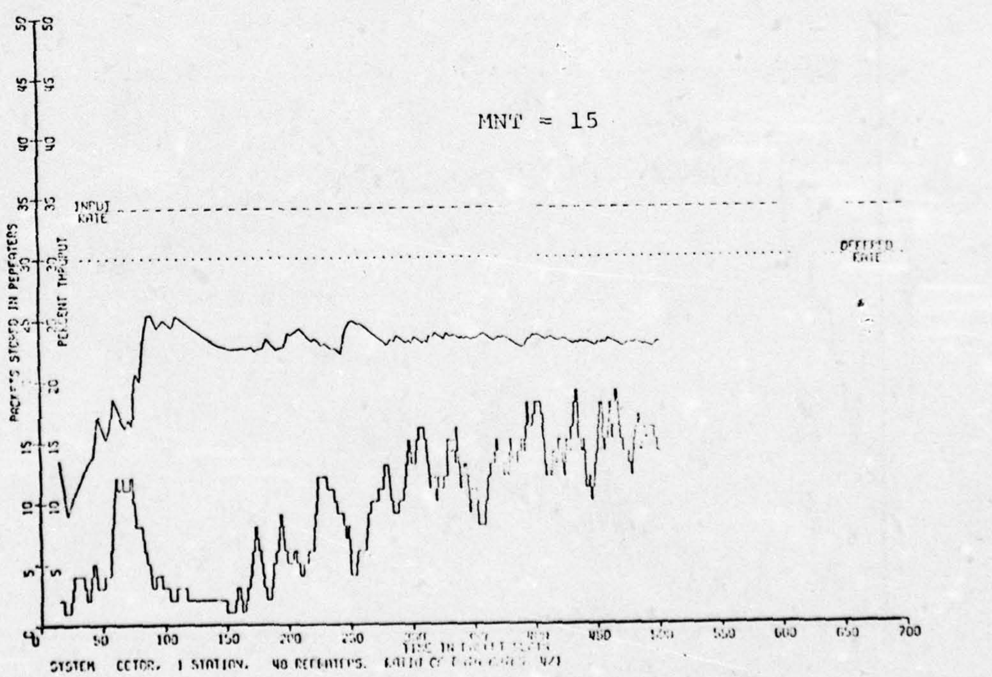
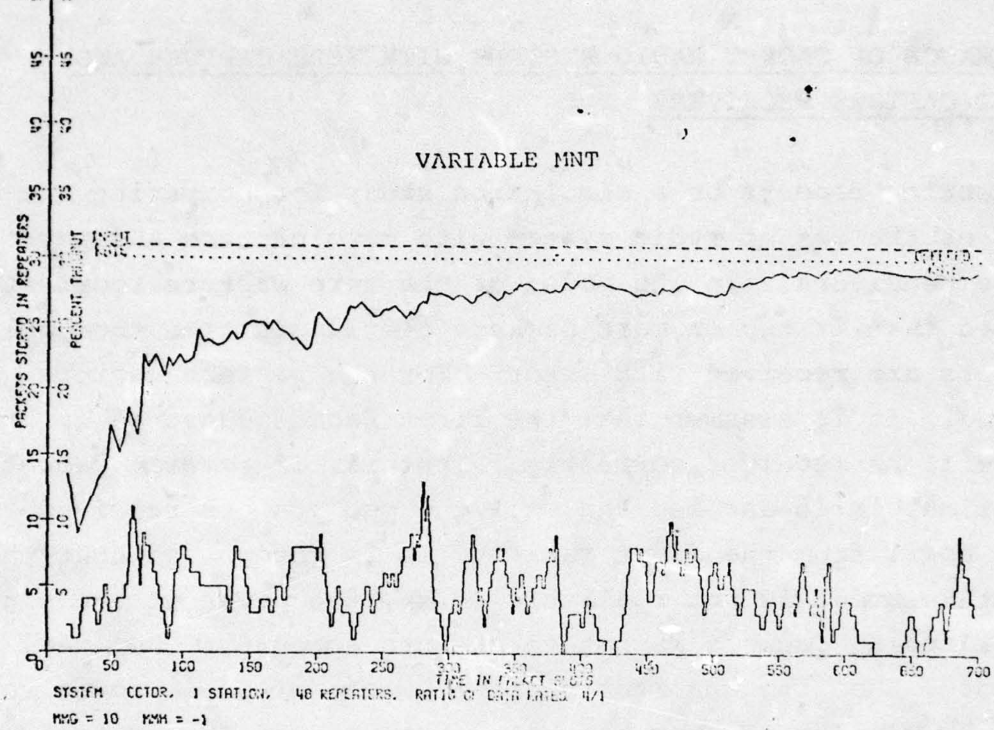


FIGURE 2.8: COMPUTER OUTPUT FOR MNT = 9 AND MNT = 15



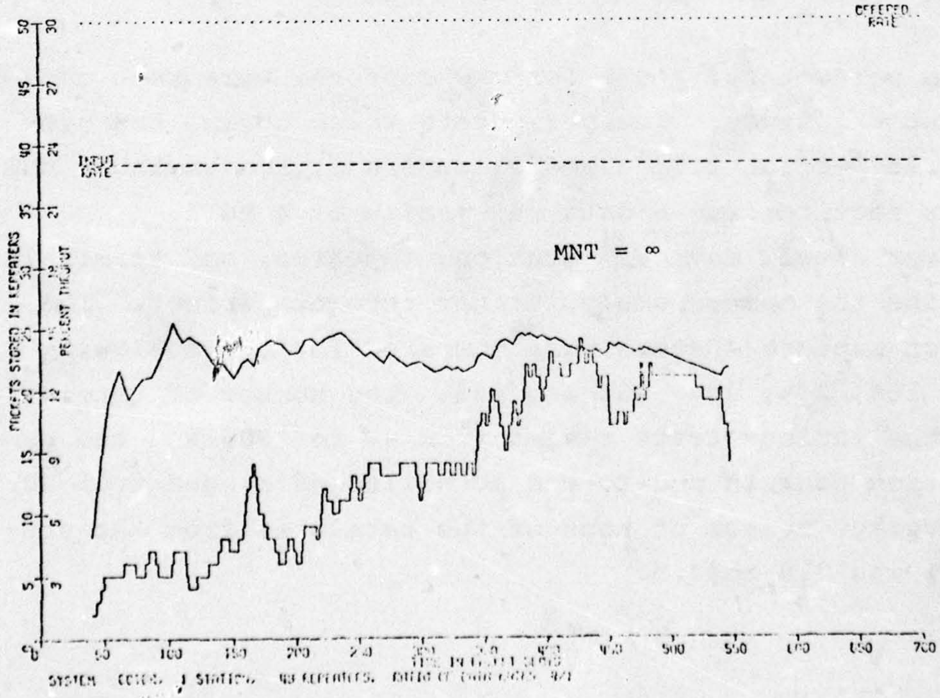


FIGURE 2.9: COMPUTER OUTPUT FOR VARIABLE MNT AND MNT = ∞

2.4 PERFORMANCE OF PACKET RADIO SYSTEMS WITH ZERO CAPTURE AND PERFECT CAPTURE RECEIVERS

This section reports on a simulation study for comparing the performance of the packet radio system with zero capture and perfect capture receivers. In the model of the zero capture receiver, it is assumed that if two or more packets overlap in time then all of the packets are received with error. For the perfect capture receiver model, it is assumed that the first packet detected by the device will be received correctly. That is, if several packets overlap in time, it is assumed that all the packets are received with error, apart from the first packet. It is recognized that the perfect capture model is not realistic (e.g., the power of the signals of overlapping packets is not taken into account). However, it is suggested that the qualitative performance of an improved receiver will be similar to what has been observed for the perfect capture model, and the quantitative results may be considered as an upper bound on the improvements in performance that one may expect.

The system parameters, apart for the capture, were kept constant throughout the study. The parameters which define the system are stated in Section 2.2, a header checksum, the variable MNT of the previous section, and a data rate ratio of 4 to 1.

The receiver of all devices, station, repeater, and terminals was assumed to be the same; namely, either zero or perfect. The zero and perfect capture systems were compared for the following offered rates: 15%, 24%, 30%, 36% and 45%. The number of terminals simulated for the various tests ranged from 40 to 100, and the number of information packets end-to-end acknowledged ranged from 90 to 300. The average number of hops of the terminals from the station (measured) was 3.9 to 4.5.

Table 2.2 summarizes the performance of the simulation runs used for comparison of the systems. Figure 2.10 shows the system throughput as a function of the traffic rate offered to the system, for the zero capture and perfect capture receivers. Figure 2.11 compares the two systems in terms of the average round trip delay, again as a function of the offered rate.

Observing the throughput curves, one can see the following two regions. When the offered rate is less than 30% (approximately), the two systems perform well and each is capable of delivering the offered traffic rate. For example, when the offered rates were 24% and 30%, the zero capture system delivered 22.46% and 29.12%, and the perfect capture system delivered 18.80% (note that the input rate was low, 19.68%) and 28.48%, respectively. Considering the same offered rate interval in the delay curves (Figure 2.11) shows that the performance of the perfect capture system is significantly better than that of the zero capture system. For the offered rates of 24% and 36%, the average round trip delays were 15.21 and 17.04 for the zero capture system, and 8.13 and 8.02 for the perfect capture system, respectively.

When the offered rate is greater than 30%, one can see from Figure 2.10 that the throughput of the zero capture system decreases, whereas that of the perfect capture system increases (although at a slower rate). Hence, the performance of the system with perfect capture receivers is superior to that with zero capture receivers both in throughput and delay for "high" offered rates. In Figure 2.11, one can see that the difference in delays between the two systems increases as a function of the offered rate.

Figure 2.12 shows the ratio of throughput to input rate as a function of the offered rate. This measure can be used to demonstrate the rate of performance degradation. It can be seen that

the perfect capture system degrades "gracefully" whereas in the zero capture system there is a relatively fast degradation above the offered rate at which its maximum throughput is obtained. This essentially implies that a system with zero capture receivers must have more sophisticated stability control procedures and that it should be designed to operate below its maximum throughput.

As a final comparison note, the self-regulation of the input rate is considered. From Table 2.2, one can see that when the offered rate is above 30%, the difference between the offered rate and the input rate for the zero capture system increases, whereas for the perfect capture system there is no significant trend. This seems to indicate that the zero capture system has an inherent control over the input rate and would tend to block terminals when the system is <u>locally</u> overloaded.

Figure 2.13 shows the average buffer occupancy in the entire repeater network. There are no significant differances and no conclusions are made. Figures 2.14 through 2.18 show the throughput and total number of packets stored in repeaters as a function of time for the various offered rates.

	OFFERED RATE [%]	INPUT RATE [%]	THROUGH-PUT [%]	TOTAL LOSS [%]	ROUND TRIP DELAY AVERAGE [SLOTS]	ROUND TRIP DELAY AVERAGE/HOP [SLOTS]	AVERAGE BUFFER OCCUPANCY [PACKETS]
ZERO CAPTURE	15	15.17	14.75	0	8.18	1.98	2.40
	24	22.97	22.46	.33	15.21	3.47	4.63
	30	30.96	29.12	2.40	17.04	3.93	5.16
	36	32.31	22.37	18.81	26.80	6.52	8.75
	45	38.92	21.05	16.98	33.96	7.63	11.50
PERFECT CAPTURE	15	15.70	15.44	0	6.15	1.50	2.95
	24	19.68	18.80	0	8.13	2.07	3.42
	30	30.07	28.48	2.22	8.02	1.88	4.82
	36	35.26	31.66	7.87	10.52	2.59	6.72
	45	45.18	32.66	1.65	17.84	4.15	10.56

TABLE 2.2: SUMMARY OF PERFORMANCE MEASURES FOR THE STUDY COMPARING ZERO VS. PERFECT CAPTURE RECEIVERS

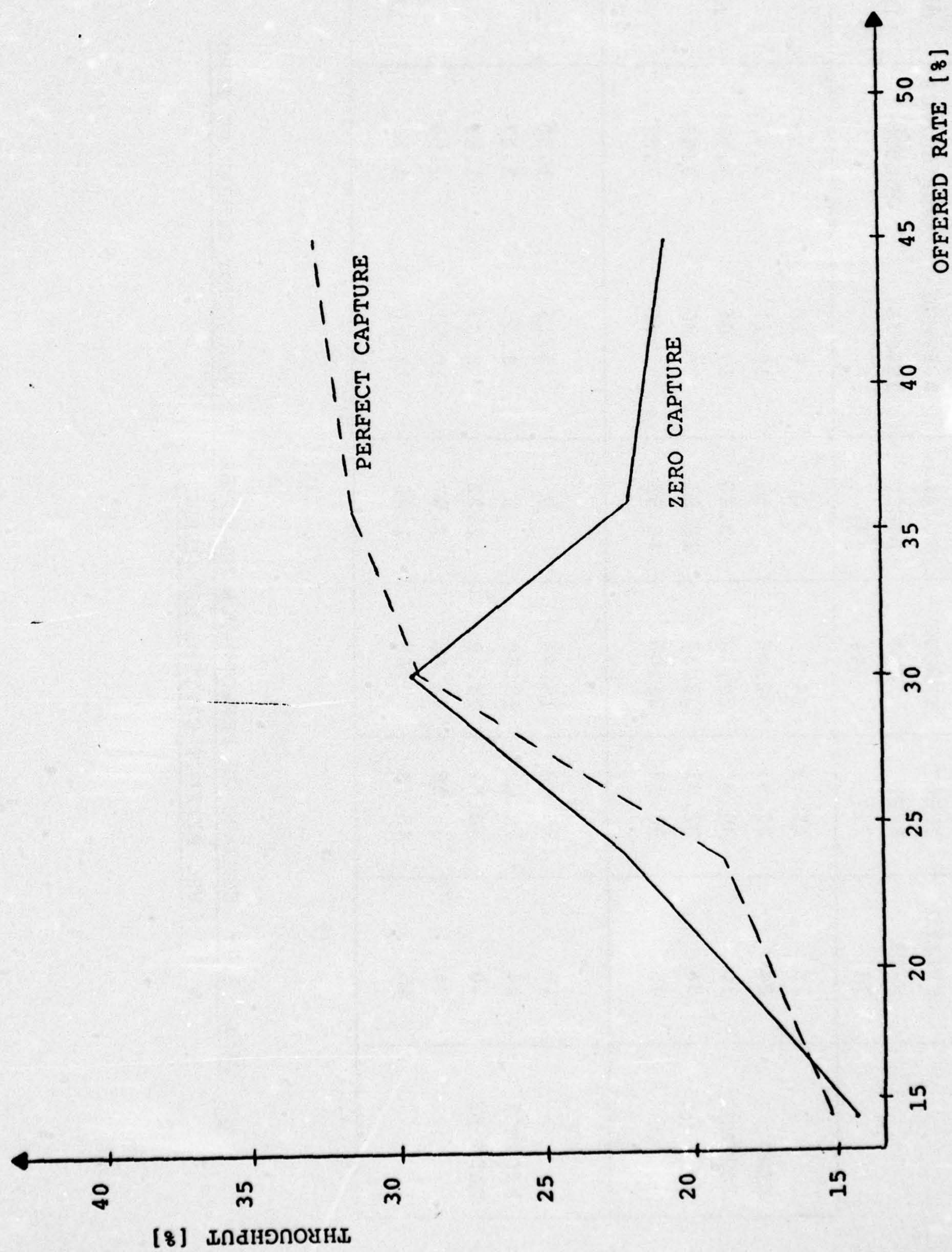


FIGURE 2.10: THROUGHPUT VS. OFFERED RATE

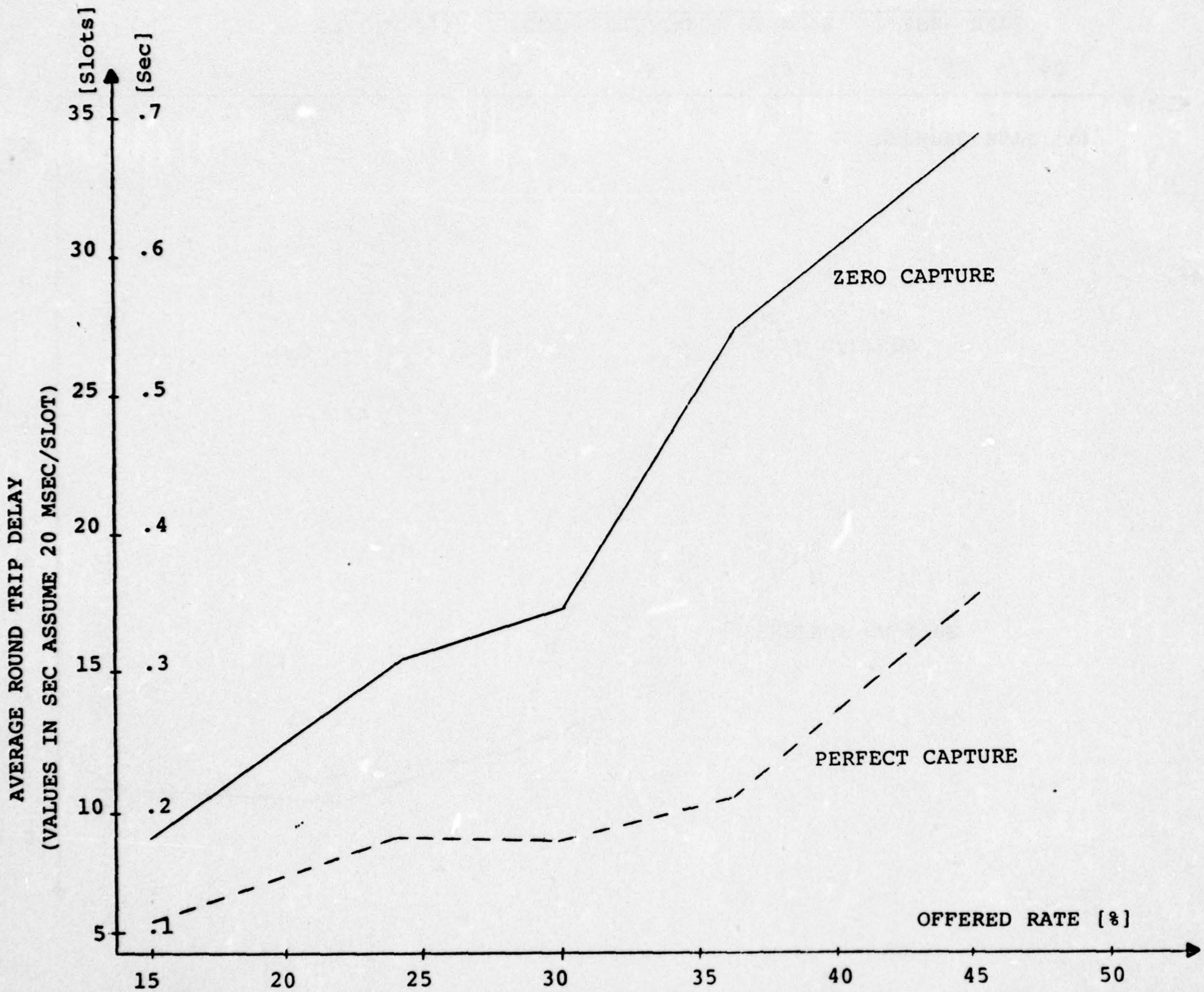


FIGURE 2.11: AVERAGE ROUND TRIP DELAY VS. OFFERED RATE

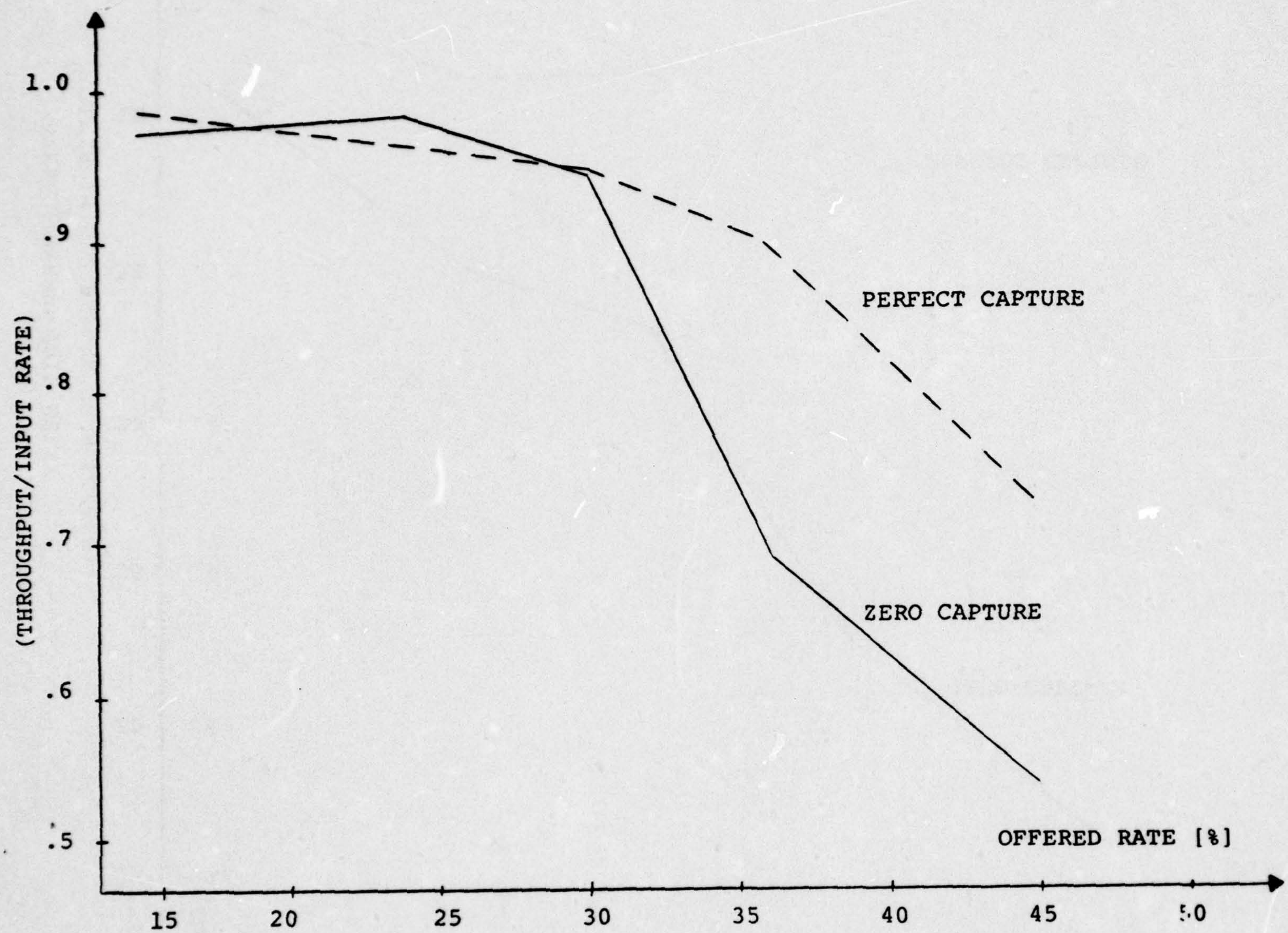


FIGURE 2.12: THROUGHPUT/INPUT RATE VS. OFFERED RATE

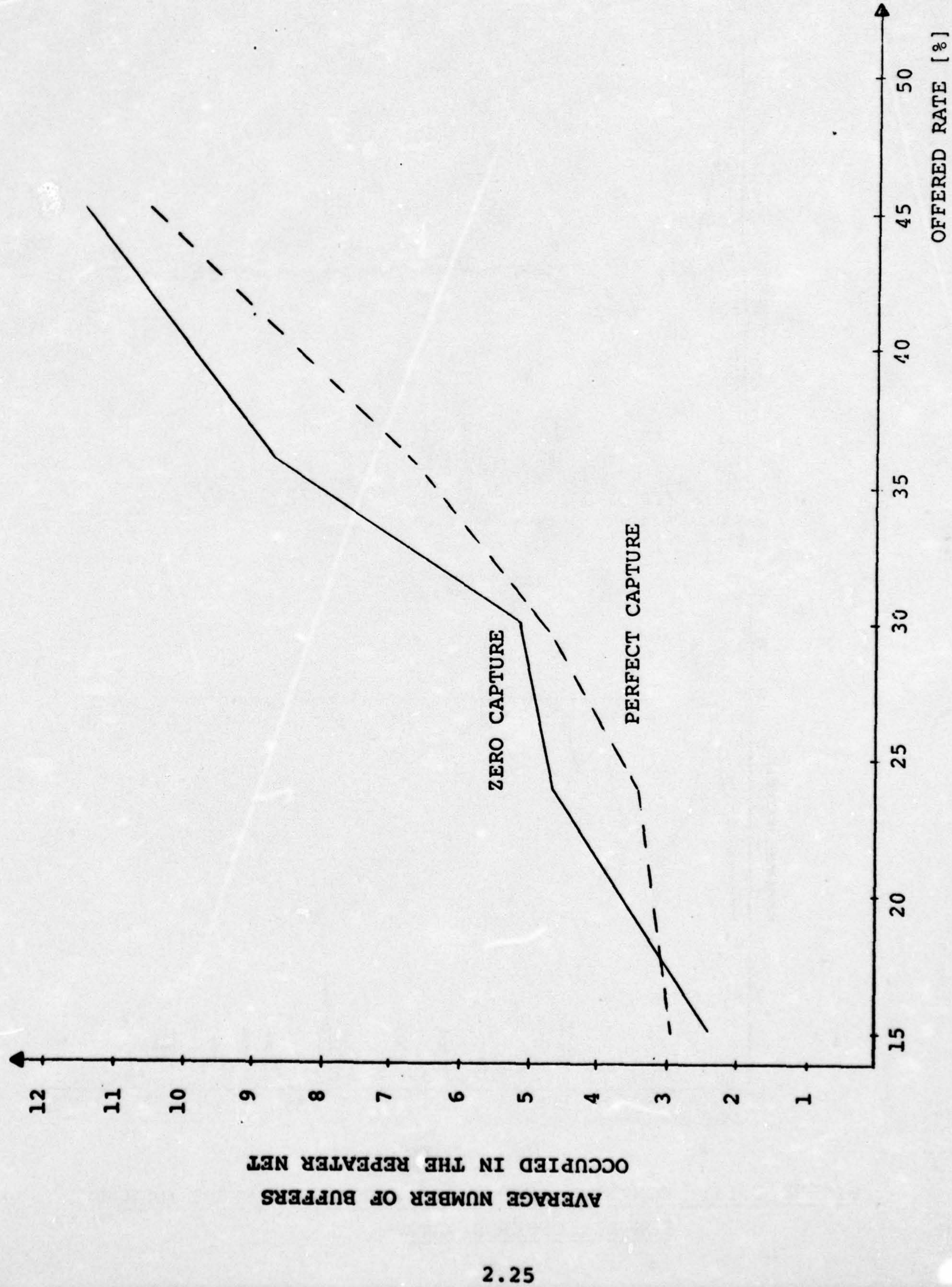
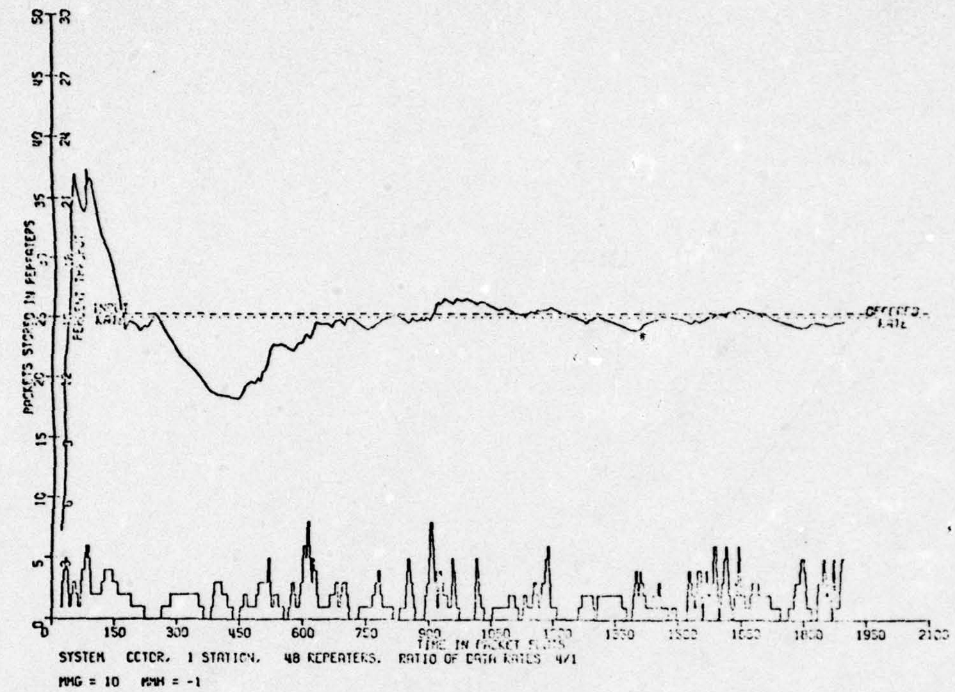
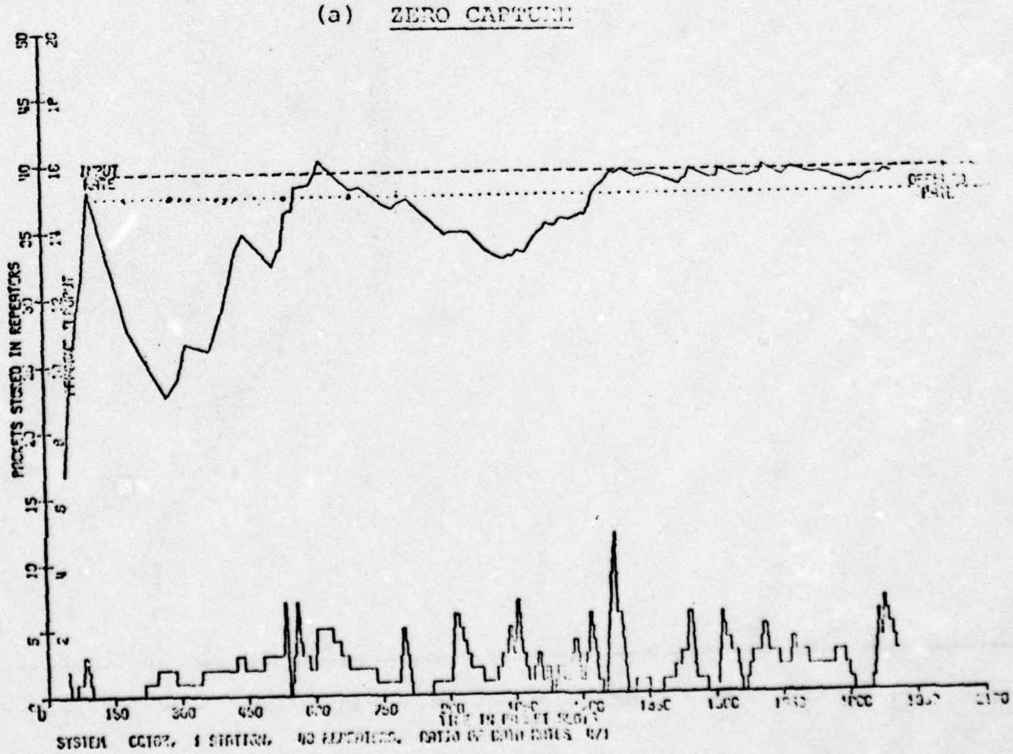


FIGURE 2.13: AVERAGE NUMBER OF BUFFERS OCCUPIED IN REPEATER NET VS. OFFERED RATE

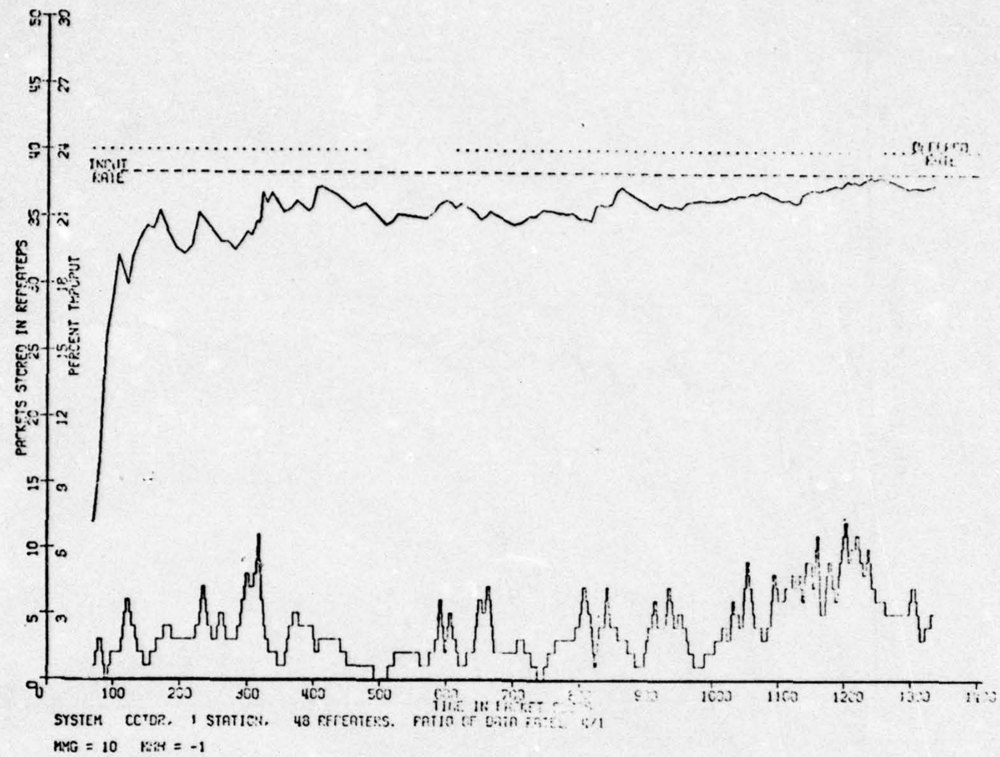


(a) ZERO CAPTURE

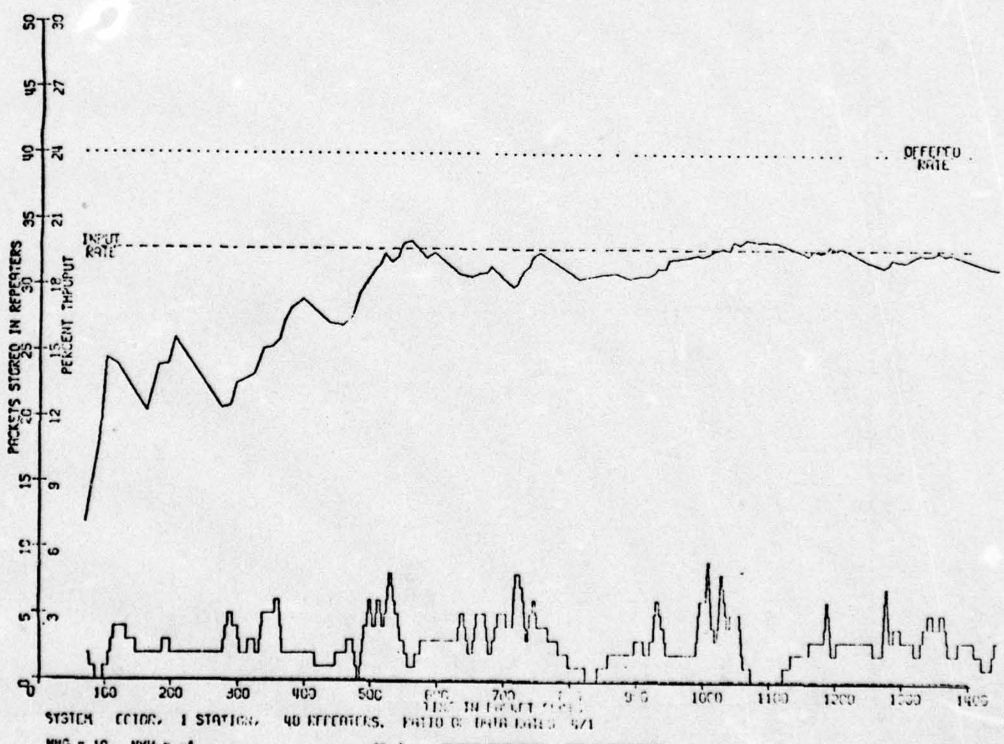


(b) PERFECT CAPTURE

FIGURE 2.14: COMPUTER OUTPUT FOR ZERO AND PERFECT CAPTURE FOR 15% OFFERED RATE

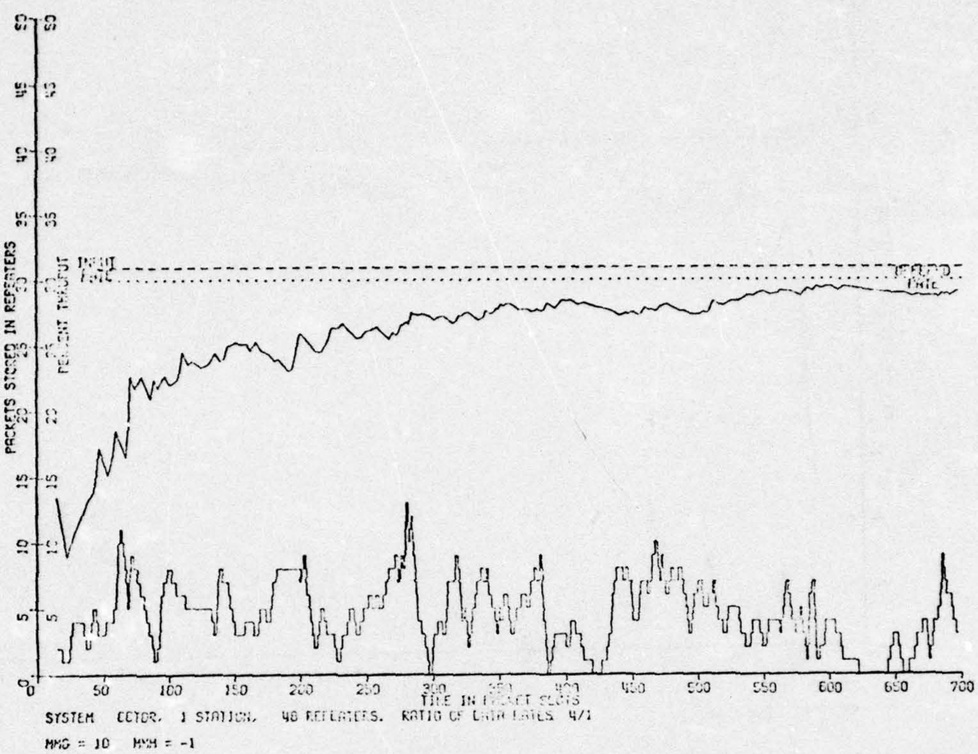


(a) ZERO CAPTURE

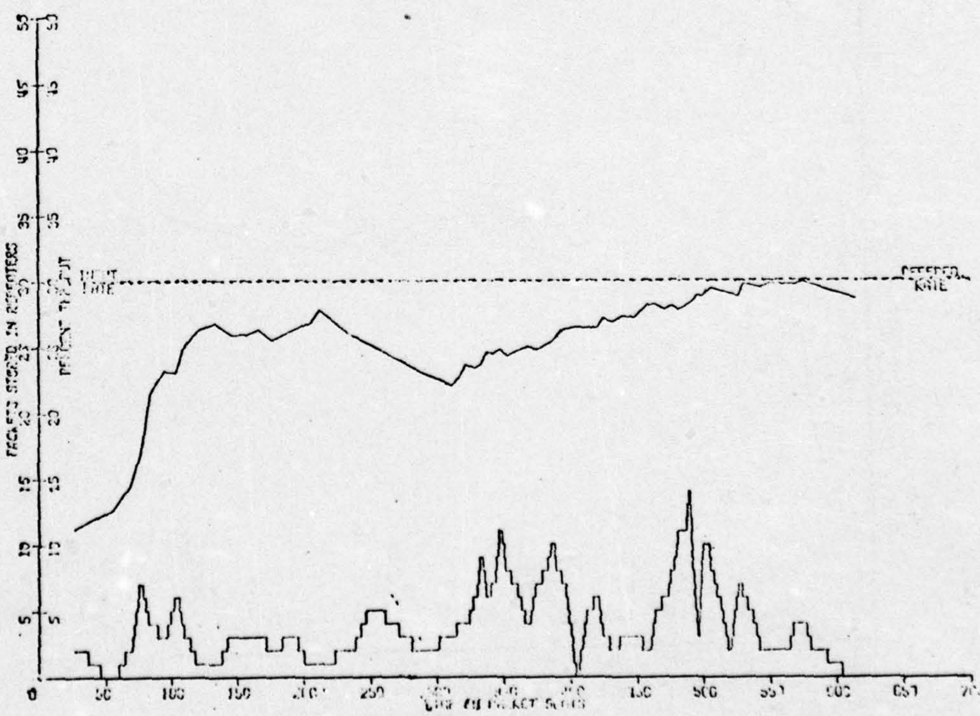


(b) PERFECT CAPTURE

FIGURE 2.15: COMPUTER OUTPUT FOR ZERO AND PERFECT CAPTURE FOR 24% OFFERED RATE

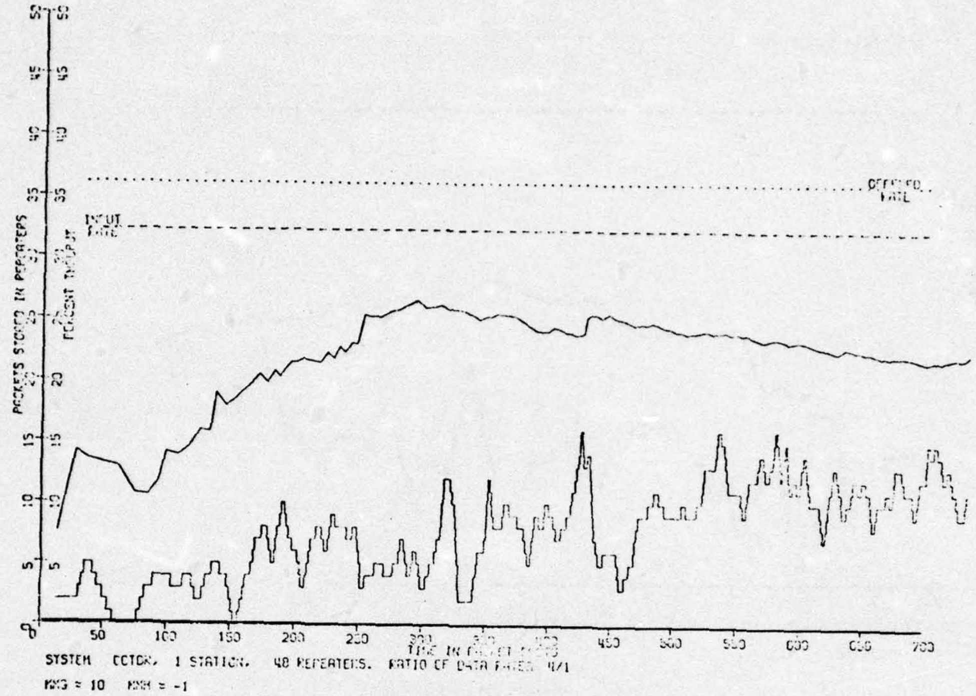


(a) ZERO CAPTURE

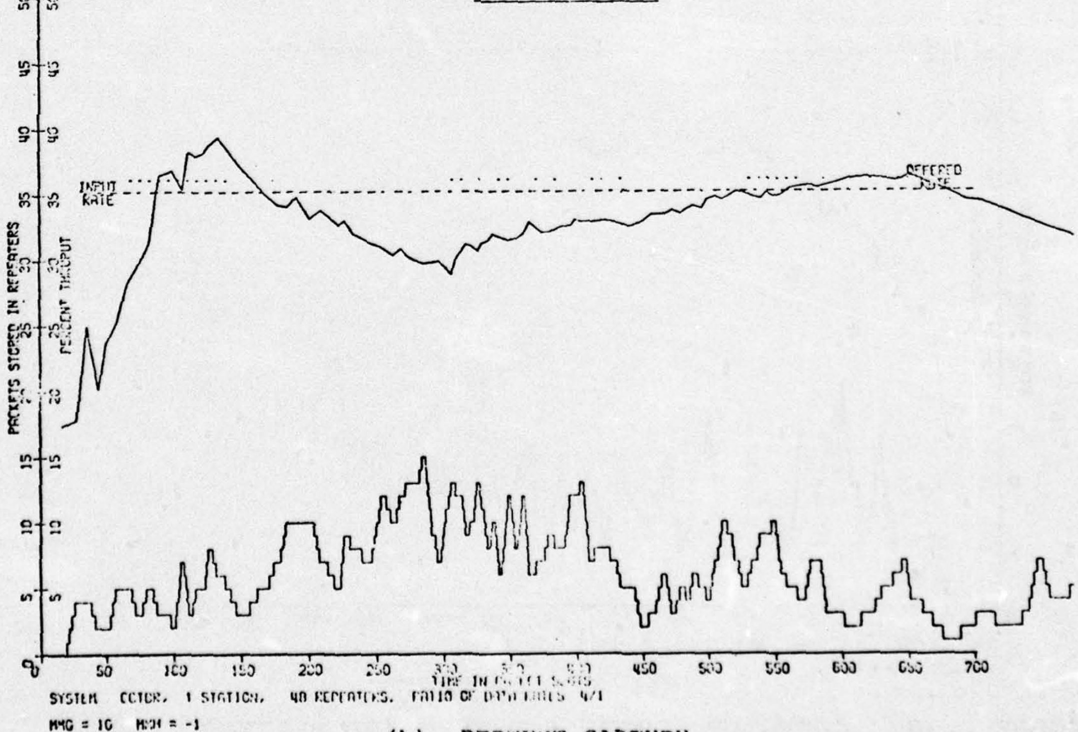


(b) PERFECT CAPTURE

FIGURE 2.16: COMPUTER OUTPUT FOR ZERO AND PERFECT CAPTURE FOR 30% OFFERED RATE

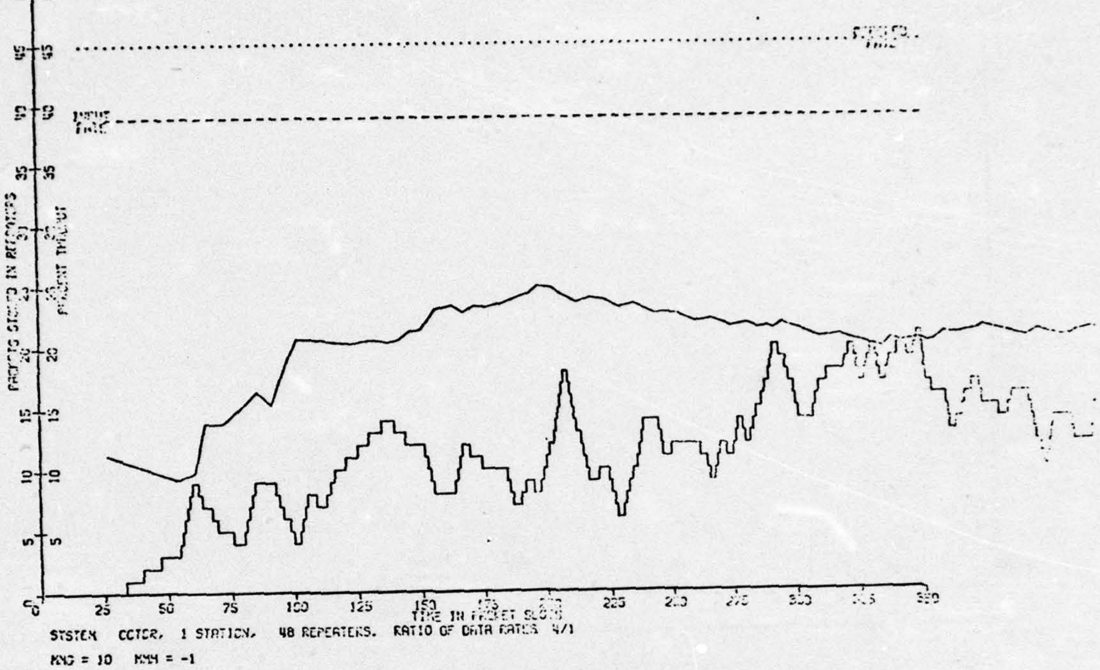


(a) ZERO CAPTURE

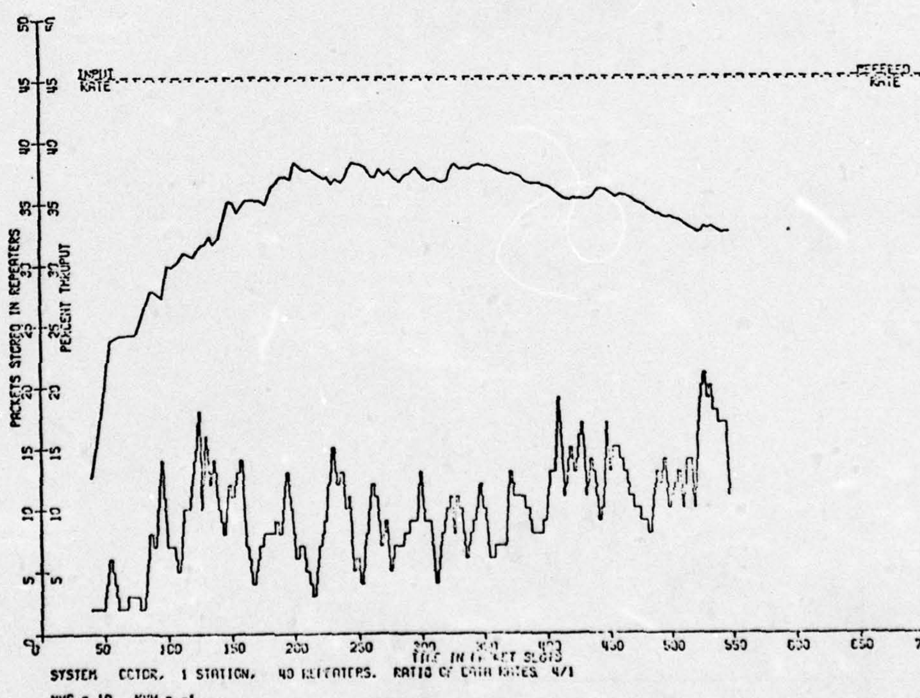


(b) PERFECT CAPTURE

FIGURE 2.17: COMPUTER OUTPUT FOR ZERO AND PERFECT CAPTURE FOR 36% OFFERED RATE



(a) ZERO CAPTURE



(b) PERFECT CAPTURE

FIGURE 2.18: COMPUTER OUTPUT FOR ZERO AND PERFECT CAPTURE FOR 45% OFFERED RATE

2.5 COMPARISON OF PACKET RADIO SYSTEMS WITH HOP-BY-HOP ACKS BASED ON HEADER AND PACKET CHECKSUMS

The modelling by simulation of the two alternative schemes for Hop-by-Hop Acknowledgments (HBH Acks) concerns the time at which the receiving device (station, repeater, terminal) examines the Header content, and hence decides whether a HBH Ack has been received. A packet format with two separate checksums, one for the header and one for the entire packet, was stipulated in the early stages of the packet radio project. This was in part to conform with a similar implementation done in the ALOHA System [ABRAMSON, 1970]. The ALOHA System is a single hop system and the objective of the header checksum was to enable the identification of the terminal which sent a packet in case the complete packet was received in error. In the packet radio system, the header checksum is utilized for a different purpose, namely to determine the HBH Ack.

Apart from the checksum the systems were identical as defined in Section 2.2, with zero capture receivers and the variable MNT scheme. The systems were compared for offered rates of 15%, 24%, 30%, and 36%. Table 2.3 summarizes the performance measures for the two systems, and Figures 2.19 through 2.22 compare the systems in terms of throughput, delay, throughput to input rate ratio, and buffer occupancy, all as a function of the offered rate.

It is evident from the figures as anticipated, that the system with the header checksum performs better than without it. The difference in performance, however, is larger than we have anticipated. From Figure 2.19 one can see that the capacity (maximum throughput) of the system without a header checksum is approximately 22% whereas that of the system with a header checksum is approximately 30%. The difference in the average round-trip delay as a function of the offered rate (Figure 2.20) is less significant. However, one should note that the delays are for different values of throughput which correspond to those shown in Figure 2.19.

It is important to note that the improvement is gained only for packets which include text. Hence the difference in performance depends on two parameters: the relative header size in an average packet and the fraction of packets which contain text. In the simulation program, there are two packet sizes: a short packet which includes a header only, and a long packet of which the header is 10% of the packet size. The fraction of header size packets in the packet radio system is not yet known. It would depend on the end-to-end protocol and upon the initialization and control schemes which will be implemented.

	OFFERED RATE [%]	INPUT RATE [%]	THROUGH-PUT [%]	TOTAL LOSS [%]	ROUND TRIP DELAY AVERAGE [SLOTS]	ROUND TRIP DELAY AVERAGE/HOP [SLOTS]	AVERAGE BUFFER OCCUPANCY [PACKETS]
HBH ACK BASED ON HEADER CHECKSUM	15	15.17	14.75	0	8.18	1.98	2.40
	24	22.97	22.46	.33	15.21	3.47	4.63
	30	30.96	29.12	2.40	17.04	3.93	5.16
	36	32.31	22.37	18.81	26.80	6.52	8.75
	45	38.92	21.05	16.98	33.96	7.63	11.50
HBH ACK BASED ON PACKET CHECKSUM	15	13.32	13.09	0	7.91	1.92	2.75
	24	21.06	19.97	0	12.49	2.96	6.89
	30	24.99	18.86	0	21.37	5.51	7.77
	36	28.23	17.91	4.00	29.76	7.10	10.02

TABLE 2.3: SUMMARY OF PERFORMANCE MEASURES FOR THE STUDY OF HEADER & PACKET CHECKSUM VS. PACKET CHECKSUM ONLY

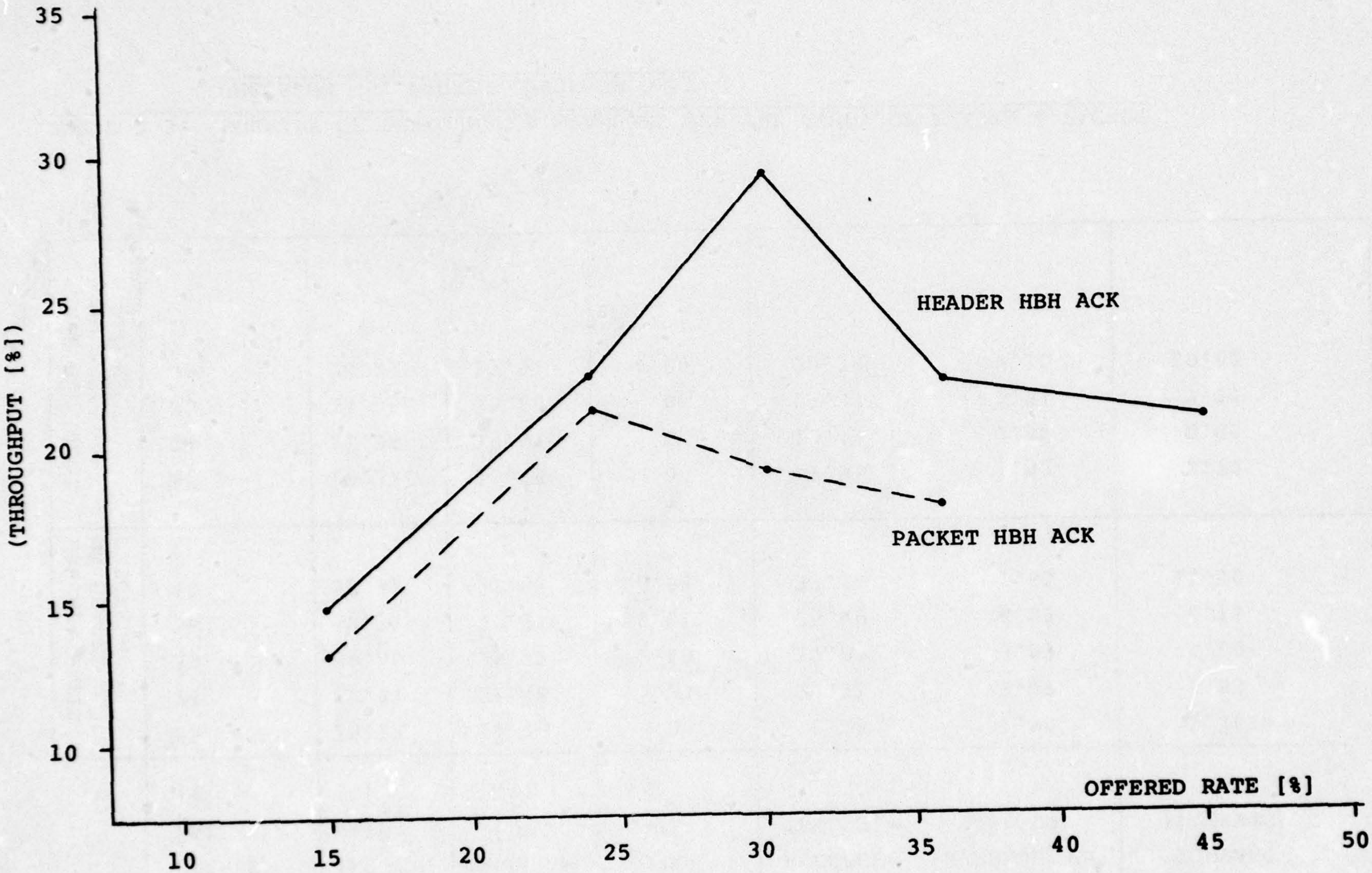


FIGURE 2.19: THROUGHPUT VS. OFFERED RATE

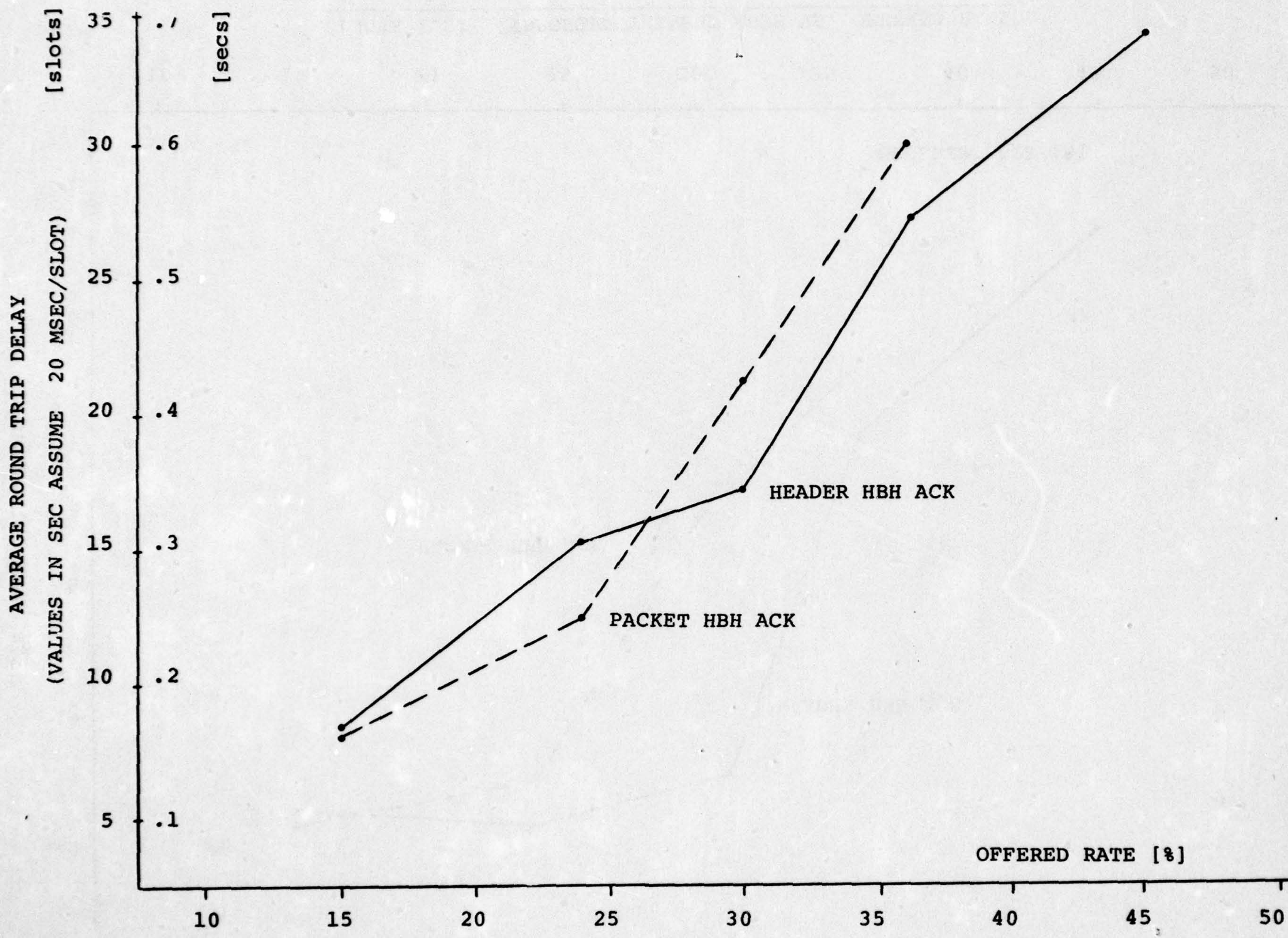


FIGURE 2.20: AVERAGE ROUND TRIP DELAY VS. OFFERED RATE

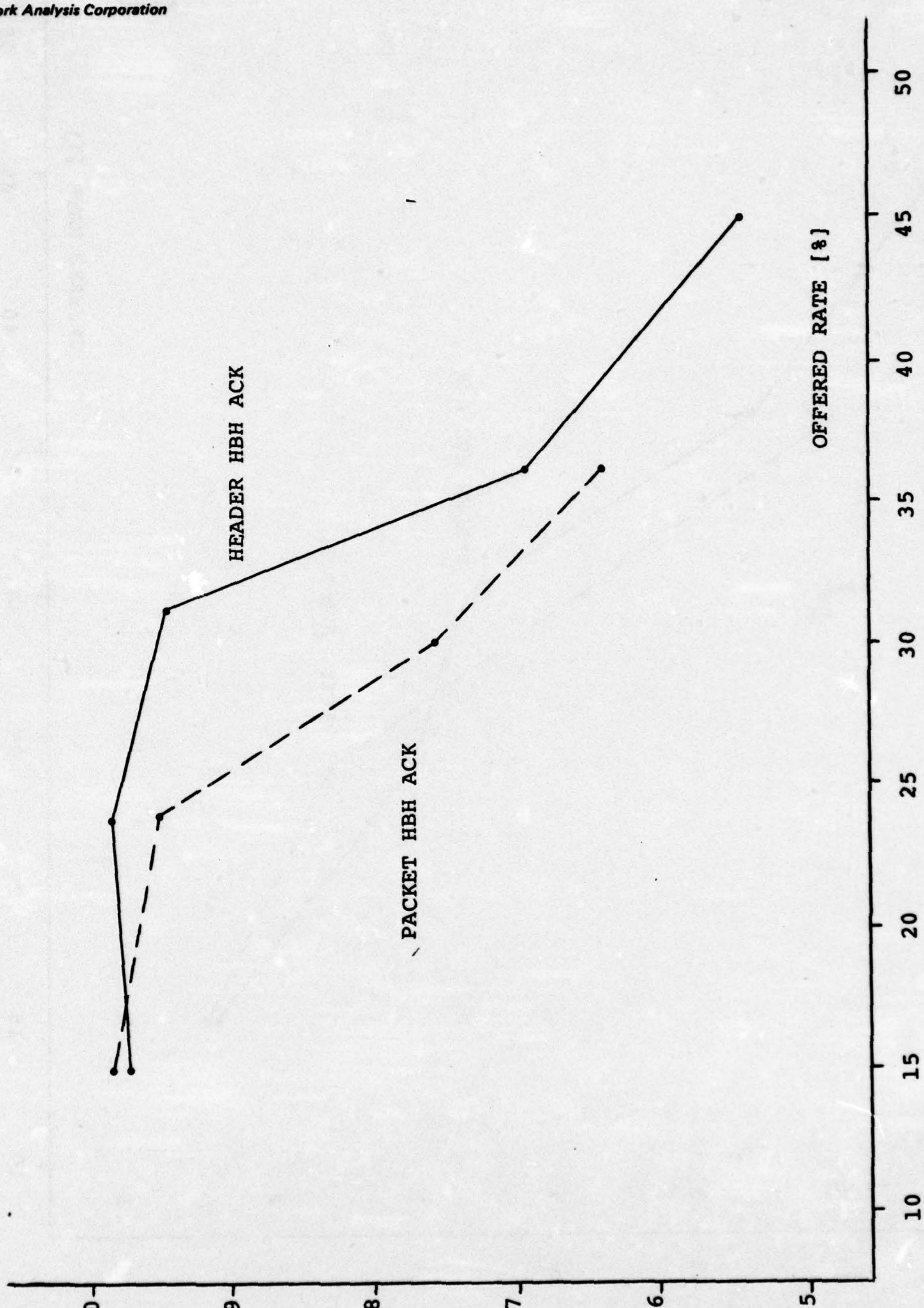


FIGURE 2.21: THROUGHPUT/INPUT RATE VS. OFFERED RATE

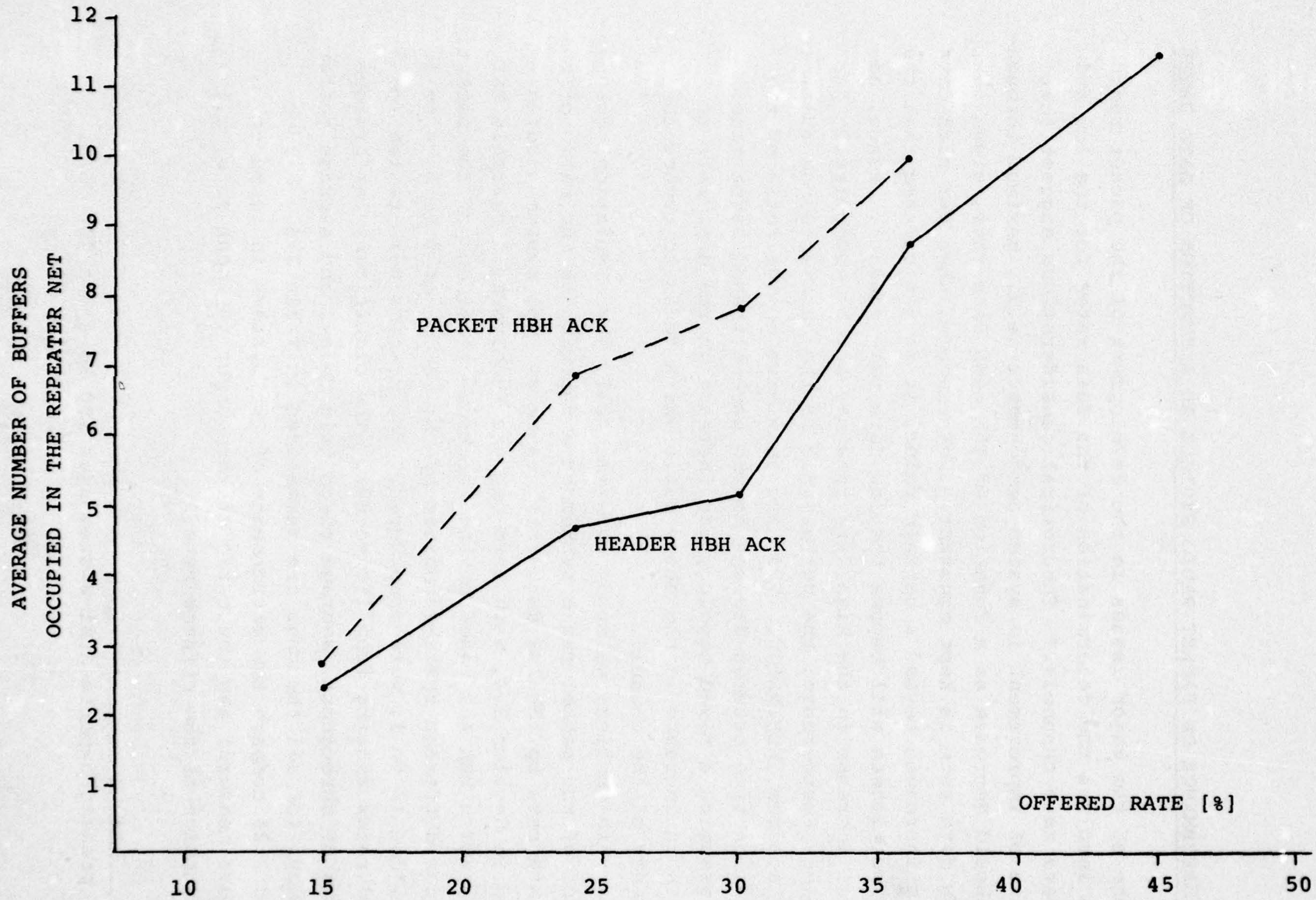


FIGURE 2.22: AVERAGE NUMBER OF BUFFERS OCCUPIED IN REPEATER NET VS. OFFERED RATE

2.6 PERFORMANCE OF PACKET RADIO SYSTEMS AS A FUNCTION OF DATA RATES

One of the major issues in the development of the packet radio system involves the determination of the data rates for the low and high data rate channels.* Theoretical considerations suggest that the rate of improvement in system performance (e.g., maximum throughput) should decrease as a function of the high data rate, given that the low data rate is kept constant. For example, when the high data rate is increased beyond a certain value, it is anticipated that the system bottleneck will become the low data rate channel. Hence, additional increase in the high data rate may not be economically justifiable. Furthermore, the performance of the carrier-sense multiple access schemes [KLEINROCK, 1975] are sensitive to the ratio of the propagation time between devices to the packet transmission time. Thus, assuming a fixed topology, an increase in the data rate results in an increase in the above ratio which tends to reduce the efficiency of the channel.

In this section, we study experimentally (by simulation) the performance of the packet radio system as a function of the ratio of the high data rate to the low data rate channels. The system studied is defined in Section 2.2, with zero capture receivers, a variable MNT scheme, and a HBH Ack based on the checksum at the end of the packet.

Four different systems defined by the ratios of 2 to 1, 4 to 1, 8 to 1, and 12 to 1, were considered. The systems were tested for offered rates ranging from 15% to 60%. The significant performance measures of throughput, average round trip delay, and average buffer occupancy, for all the runs are summarized in Table 2.4. Figure 2.23 to 2.26 compare the performance of the systems in terms of the above measures and the ratio of throughput to input rate, all as a function of the offered rate.

---

* The present experimental system uses 100 Kb/s and 400 Kb/s.

Apart from the buffer occupancy which does not show a definite trend to justify conclusions, all other measures of performance clearly demonstrate the following. An increase in the ratio from 2/1 to 4/1 significantly improves system performance. Similarly, an increase of the ratio from 4/1 to 8/1 also demonstrates a significant improvement in performance. On the other hand, the performance of the system with the ratio 12/1 is almost the same as that of the system with the ratio 8/1, hence no significant improvement is observed in this range. For example, for an offered rate of 30%, the measured throughputs are 15.69%, 18.86%, 25.15%, and 26.29% for the systems with ratios 2/1, 4/1, 8/1, and 12/1, respectively. Similarly, the measured average round trip delays for a load of 30% offered rate are 24.99, 21.37, 10.39, and 9.77 packet slots for the systems with ratios 2/1, 4/1, 8/1, and 12/1, respectively.

Assuming that the low data rate channel is 100 Kb/s, this study demonstrates that the best choice for the high data rate channel will be between 400 Kb/s and 800 Kb/s. No significant gain is obtained when the data rate is increased beyond 800 Kb/s as is clearly demonstrated in Figures 2.23 through 2.26.

OFFERED RATE [%]	THROUGHPUT [%] RATIO OF DATA RATES				AVERAGE ROUND TRIP DELAY [SLOTS] RATIO OF DATA RATES				AVERAGE BUFFER OCCUPANCY [PACKETS] RATIO OF DATA RATES			
	2/1	4/1	8/1	12/1	2/1	4/1	8/1	12/1	2/1	4/1	8/1	12/1
15	10.15	13.09			12.03	7.91			5.18	2.75		
24	14.65	19.97	22.86	24.19	13.69	12.49	11.94	8.68	6.13	6.89	5.84	4.56
30	15.69	18.86	25.15	26.29	24.99	21.37	10.39	9.77	8.24	7.77	6.55	5.10
36	13.96	17.91	26.44	28.22	23.85	29.76	21.12	19.63	7.70	10.02	7.86	9.83
45			28.70	29.31			19.68	21.28			10.00	9.60
60			26.58	28.11			22.21	21.56			9.15	9.60

TABLE 2.4: SUMMARY OF PERFORMANCE MEASURES FOR THE STUDY OF RATIO OF CHANNEL DATA RATES

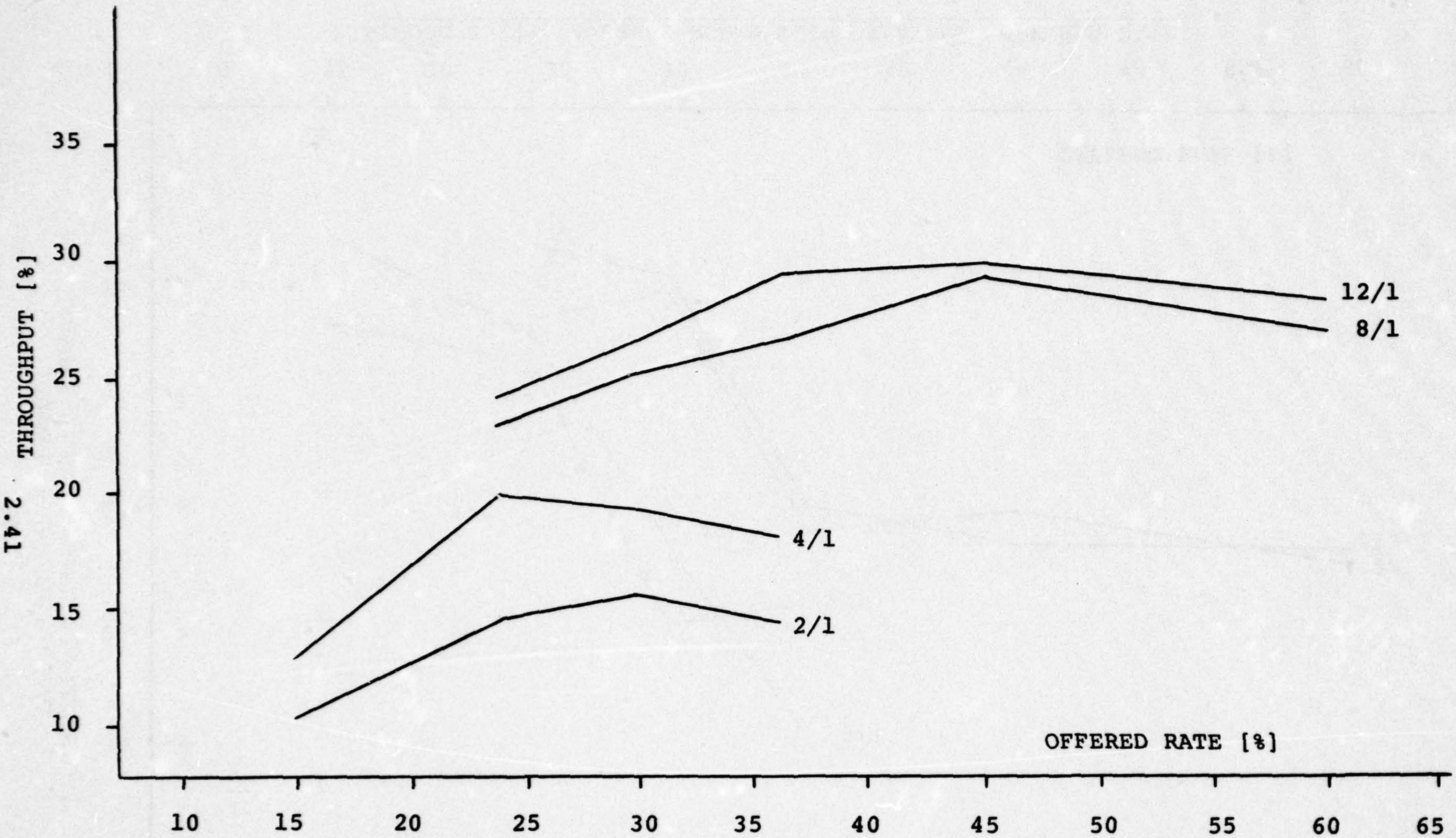


FIGURE 2.23: THROUGHPUT VS. OFFERED RATE

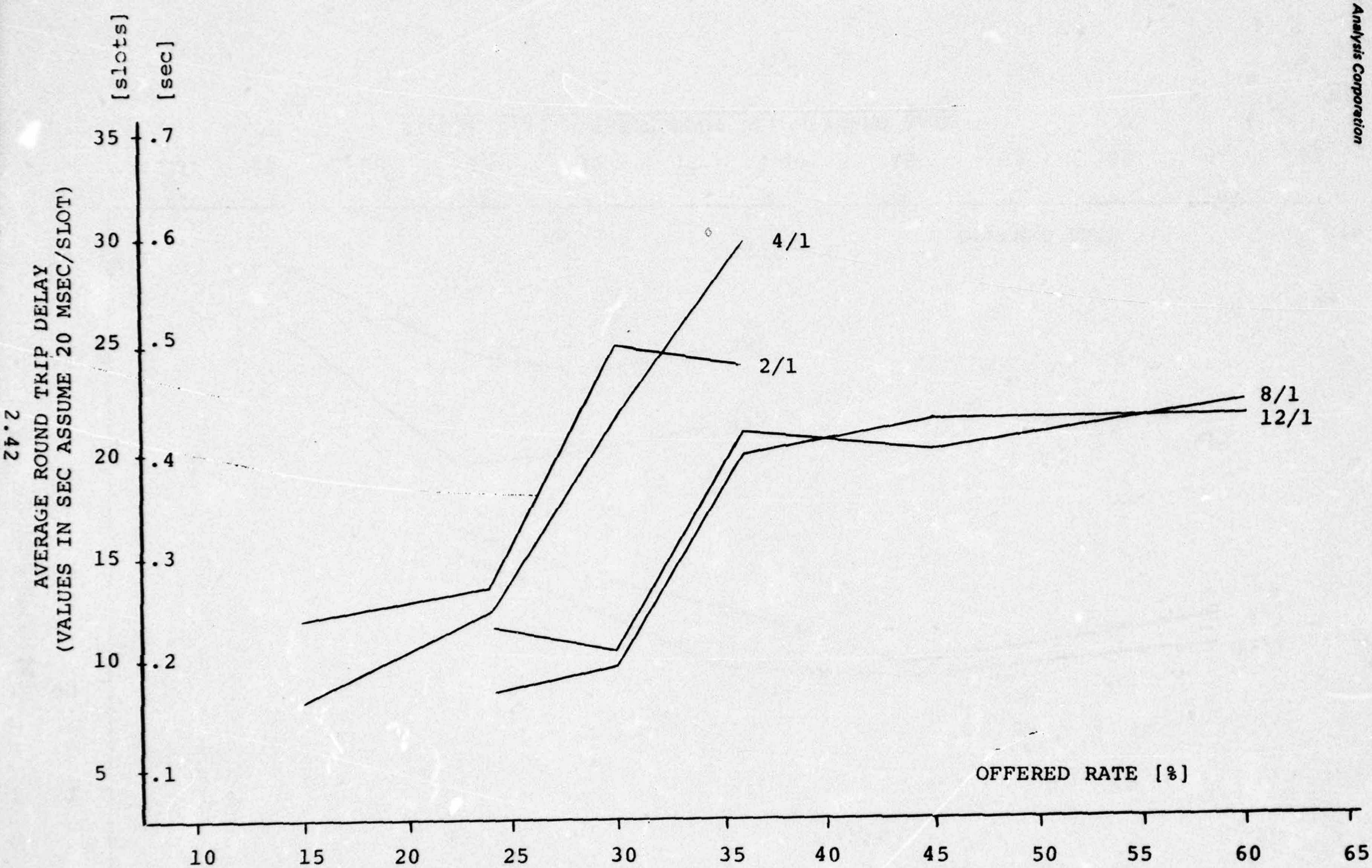


FIGURE 2.24: AVERAGE ROUND TRIP DELAY VS. OFFERED RATE

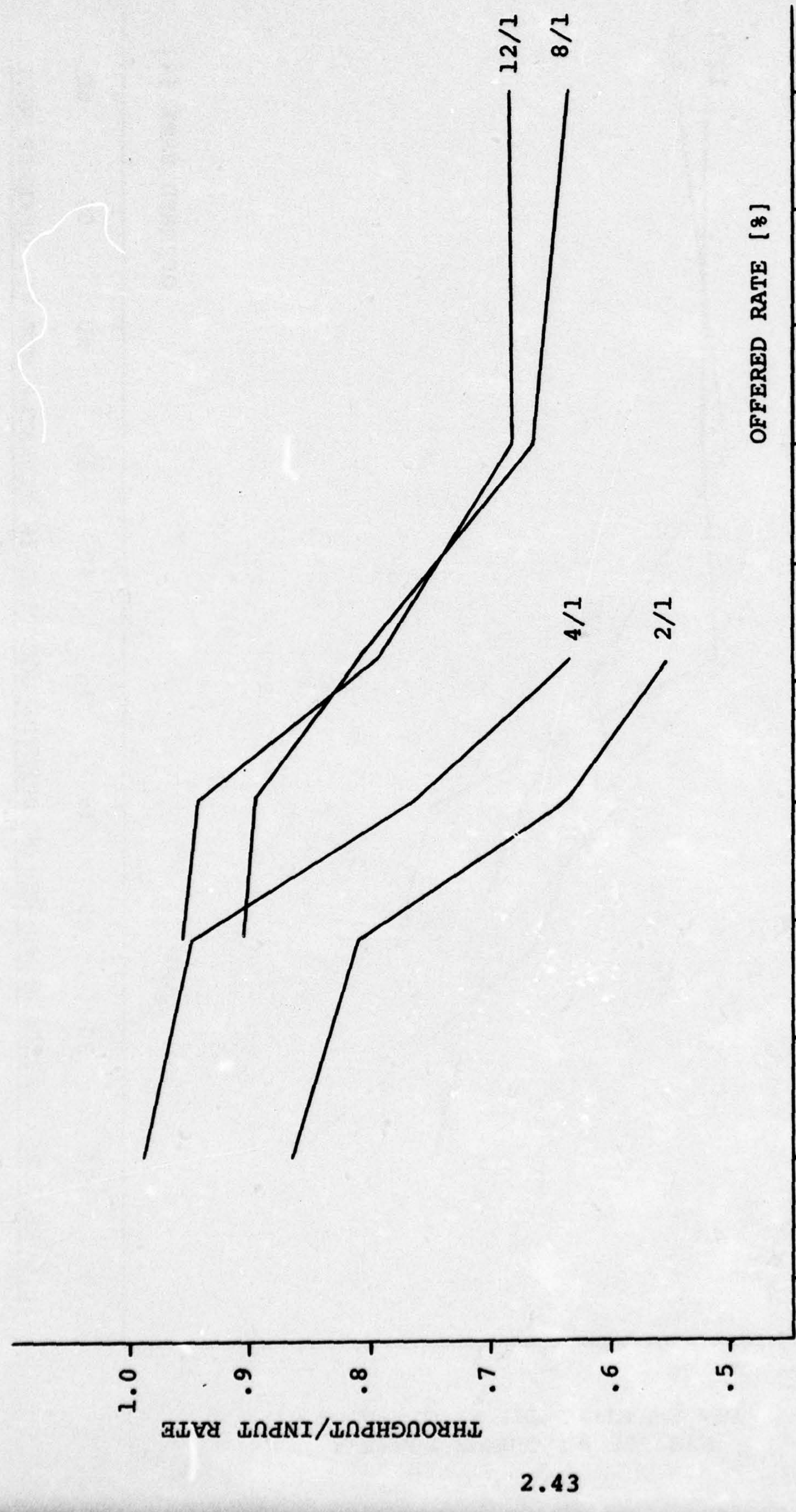


FIGURE 2.25: THROUGHPUT/INPUT RATE VS. OFFERED RATE

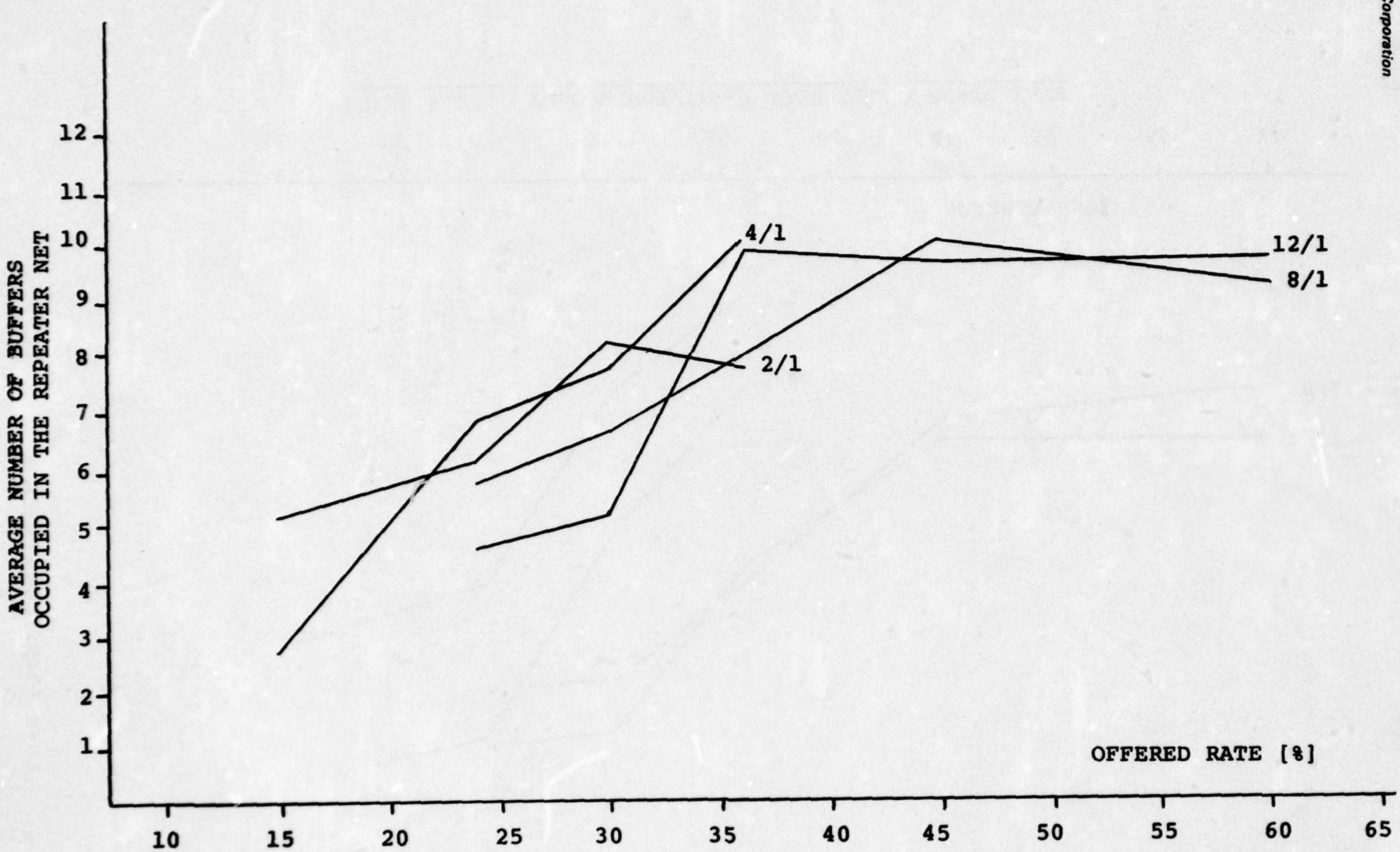


FIGURE 2.26: AVERAGE NUMBER OF BUFFERS OCCUPIED IN REPEATER NET VS. OFFERED RATE

2.7 CONCLUSIONS

Several specific design alternatives have been studied using a simulation approach. The alternative packet radio systems which result for a given design issue were compared on the basis of relative performance. The major conclusions of the study were:

1. The maximum number of packet transmissions, MNT, before discarding the packet should be software modifiable. A variable MNT as a function of the hierarchy level demonstrated the best performance. When a fixed MNT is used, a small value of MNT between 3 and 6 is preferable to large values.

2. A perfect capture receiver significantly improves system performance in two aspects. It increases the system capacity (maximum throughput) for a given value of delay, and it results in a system which is less sensitive to overload fluctuations by demonstrating gradual ("graceful") degradation as compared to the system with zero capture receivers.

3. A system which uses a header checksum, utilized for HBH Ack, performs better than one which uses only a checksum at the end of the packet. The difference in performance depends upon the relative header and packet sizes as well as the fraction of short (header size) packets in the network.

4. For a given 100 Kb/s low data rate channel, it is demonstrated that the "optimum" value for the high data rate channel will be between 400 Kb/s and 800 Kb/s. No significant gain in system performance is obtained when the high data rate channel is increased beyond 800 Kb/s.

It is noted that all the studies of this chapter were done for a single station packet radio network. It is not anticipated that any of the conclusions of the studies presented in this chapter will change for multistation packet radio systems. However, it is noted that other design issues which are specific to multistation networks need to be studied.

VOLUME 2

Chapter 2

REFERENCES

REFERENCES

[ABRAMSON, 1970] Abramson, N., "The ALOHA System - Another Alternative for Computer Communication," 1970 Fall Joint Computer Conference, AFIPS, Conference Proceedings, Vol. 37, pp. 281-285.

[FRANK, 1975] Frank, H., I. Gitman, and R. M. Van Slyke, "Packet Radio System - Network Considerations," Proceedings of the National Computer Conference, May, 1975, pp. 217-231.

[GITMAN, 1976] Gitman, I., R. M. Van Slyke, and H. Frank, "Routing in Packet Switching Broadcast Radio Networks," IEEE Trans. on Communications, August, 1976.

[KLEINROCK, 1975] Kleinrock, L., and F. Tobagi, "Random Access Techniques for Data Transmission Over Packet Switched Radio Channels," Proceedings of the National Computer Conference, May, 1975, pp. 187-201.

[NAC, 1974] (Network Analysis Corporation), "Simulation of Packet Switched Communication Networks," The Practical Impact of Recent Computer Advances on the Analysis and Design of Large Scale Networks, Third Semiannual Technical Report, June, 1974, available from Defense Communication Center, Arlington, Virginia.

REFERENCES (Cont'd.)

[NAC, 1975a] (Network Analysis Corporation), "Maximum Number of Transmissions for Packet Radio Repeaters and Stations," Packet Radio Temporary Node #153, available from ARPA Network Information Center, Stanford Research Institute, Menlo Park, California.

[NAC, 1975b) (Network Analysis Corporation), "Comparison of Packet Radio Networks with Zero Capture and Perfect Capture Receivers," Packet Radio Temporary Note #154, available from ARPA Network Information Center, Stanford Research Institute, Menlo Park, California.

VOLUME 2

Chapter 3

STABILITY CONSIDERATIONS IN PACKET RADIO NETWORKS

3. STABILITY CONSIDERATIONS IN PACKET RADIO NETWORKS

3.1 GENERAL

Packet communication systems with random access to a common broadcast channel may, under unfavorable traffic load conditions, become very inefficiently utilized (or even unstable) if proper controls are not implemented [LAM, 1974]. For example, Figure 3.1 shows a typical behavior of effective throughput S versus input transmission rate G (including retransmissions) for a one hop random access system. There is a value G* which maximizes throughput. For G G* throughput performance degrades because of increasing channel interference. Thus, a direct (or indirect) control on G must be applied.

Also shown in Figure 3.1 is a typical load curve representing user throughput requirements as a function of transmission rate G [LAM, 1974]. For a system with a finite number of users the offered load typically decreases with G, since higher values of G imply higher number of retransmissions per packet, higher end-to-end delay, and consequently a reduction in new packet input rate. The intersections of system curve and load curve correspond to possible equilibrium points [LAM, 1974]. For the situation displayed in Figure 3.1, we have three equilibrium points A, B, and C. Only A and C however, are stable, while B is a point of transient equilibrium. Clearly, the operation at point C, although stable, is not very efficient. Therefore, A is the desirable equilibrium point in normal operating conditions. Since load fluctuations may cause the equilibrium to shift from A to C, we need stability control procedures to maintain the equilibrium at A, and recovery procedures to recover from degraded mode C.

Much work has been done to characterize stability and to develop control procedures for two specific random access systems, namely the ALOHA systems and the satellite broadcast system

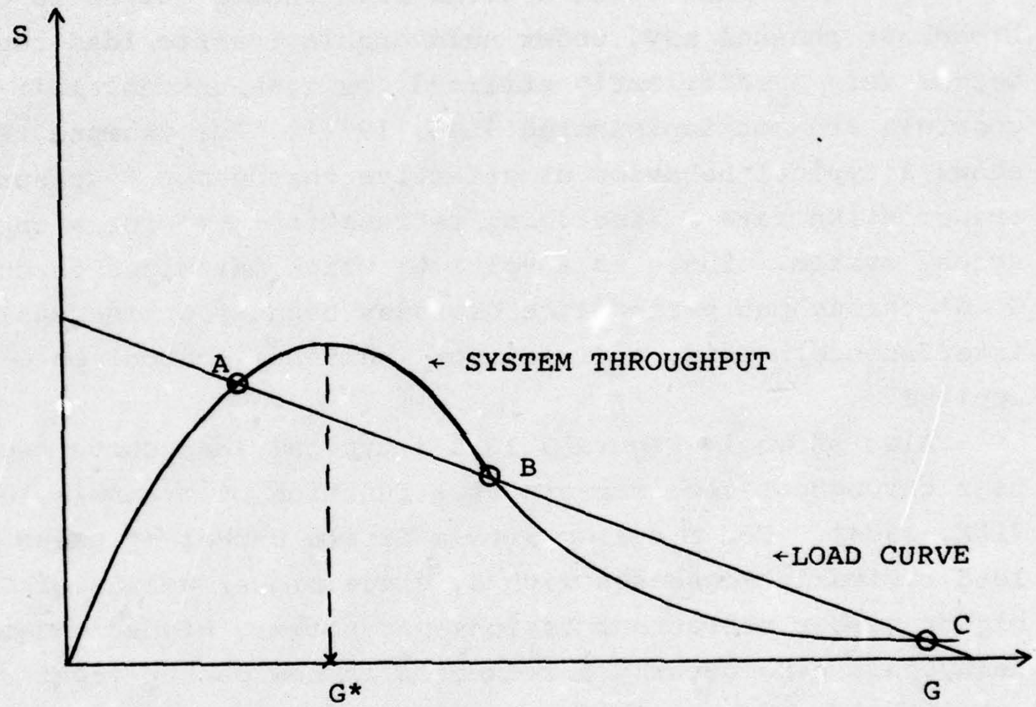


FIGURE 3.1: TYPICAL THROUGHPUT VS. INPUT TRANSMISSION RATE BEHAVIOR FOR A RANDOM ACCESS SYSTEM

[FERGUSON, 1975], [LAM, 1974]. The results have recently been extended to a simple Packet Radio (PR) model [TOBAGI, 1975]. Generally, the proposed stability procedures consist of the dynamic regulation of user retransmission intervals based on local and/or global measurements.

The above models and procedures, however, are limited to systems with only one hop and with traffic only in one direction (from terminals to station, or to satellite). In a general PR network (PRNET) configuration, with several levels of repeaters, a large terminal population, and two-way traffic from and to station, the stability control problem is far more complex. In fact, a variety of design parameters must be considered which were not present in the one-hop, one-way systems, namely number and location of repeaters, power range and interference between radio devices, retransmission rates of station and repeaters, maximum acceptance rate of packets at repeaters, repeater buffer storage, etc. None of the above parameters are under user control (as the retransmission parameters were for the ALOHA or the satellite channel) and therefore must be optimized in the design phase, or dynamically adjusted during network operation by internal control procedures. In addition, there are end-to-end protocol parameters that are under a more direct control of the user, namely transmission and retransmission rates from terminals, maximum number of outstanding packets allowed, retransmission time outs, etc.

Here we propose a two-phase approach to identify and optimize the critical stability parameters in a PRNET.

In Phase I, we study the behavior of the system under the assumption that user terminals transmit with steady rate G (in reality G is not an independent variable, but will depend on user behavior, flow control procedures, end-to-end delays, etc.). In particular we evaluate network throughput S, the sum of the inbound and outbound traffic, as a function of G for various network parameters. We then adjust network parameters and develop control

procedures so as to best protect the network from congestion (or, better, to optimize the trade off between stability and other performance measures such as delay, bandwidth, etc.).

Phase II of the stability study is concerned with user behavior and the impact of user protocols on network stability. We investigate control procedures such as dynamic adjustment of the retransmission time out as a function of measured end-to-end delay, adjustment of outstanding packet window, limitation on the number of terminals simultaneously logged-in at the station, etc. The evaluation and "tuning" of user control procedures is performed assuming that the PRNET internal parameters have been previously optimized during Phase I.

Basically, our approach separates the optimization of internal network parameters from the optimization of user related parameters in order to simplify the problem. The two optimizations, however, are interrelated and would probably lead to more effective results if carried out simultaneously. For example, if the majority of the user population consists of computers or intelligent terminals which support sophisticated end-to-end protocols, it may be cost-effective to rely more on user procedures rather than on internal PRNET stability procedures. If, on the other hand, the user population consists of non sophisticated terminals (e.g., sensors, teletypes, etc.), a user-independent, network based stability control procedure is required. In addition, other considerations such as access priorities for terminals to repeaters, fair allocation of network resources to users (regardless of their hop distance from the station) etc., may require control procedures which should be independent from user protocols.

In general, the best combination between user based and network based stability procedures will depend on user requirements and characteristics, and a final tuning of the stability strategy is required for each application.

The purpose of this study is to gain an understanding of network performance and stability as a function of network parameters, and to develop network procedures to improve stability. We focus on network based procedures, and leave user protocols for a later discussion.

Various models for 1-hop, 2-hop and multihop PRNET's are defined and analyzed. Due to the complexity of the models, simplifying assumptions are made whenever necessary to obtain approximate closed form solutions. The emphasis is on the determination of general performance-stability trends rather than on the rigorous analytical investigation of the models.

Based on approximate analytical results, we then show that network stability is critically related to the behavior of the hop between first level repeaters and the station. In particular, transmission rates of station and first level repeaters must be properly balanced to optimize network efficiency and avoid station buffer overflow. Furthermore, the repeater to station hop is typically the bottleneck of the entire system. Therefore, terminal and station input rates must be regulated in order to prevent overload of this hop.

Based on the above properties, two control procedures are proposed for the improvement of network stability:

1. Control of hop retransmission intervals for station and repeaters (to balance the respective retransmission rates).

2. Control of input rate of terminal packets into the network.

Simulation experiments are being carried out to evaluate the performance of the control procedures.

3.2 ANALYTICAL MODELS

In this section, we study various analytical models for Packet Radio networks. For each model, the analytical expression of network throughput is derived as a function of offered rate and other system parameters. Stability behavior and bottleneck properties are discussed.

3.2.1 One-Hop, Infinite Terminal Population

Model description (see Figure 3.2):

- Single station.
- Infinite terminal population.
- Slotted ALOHA channel.
- Infinite buffer storage (in the station).

The steady state equations for the model are summarized below [GITMAN, 1974]:

$$\begin{cases} S_s = G_s e^{-G_1} \\ S_1 = G_1 e^{-G_1}(1-G_s) \\ S_s = \alpha S_1 \end{cases} \tag{3.1}$$

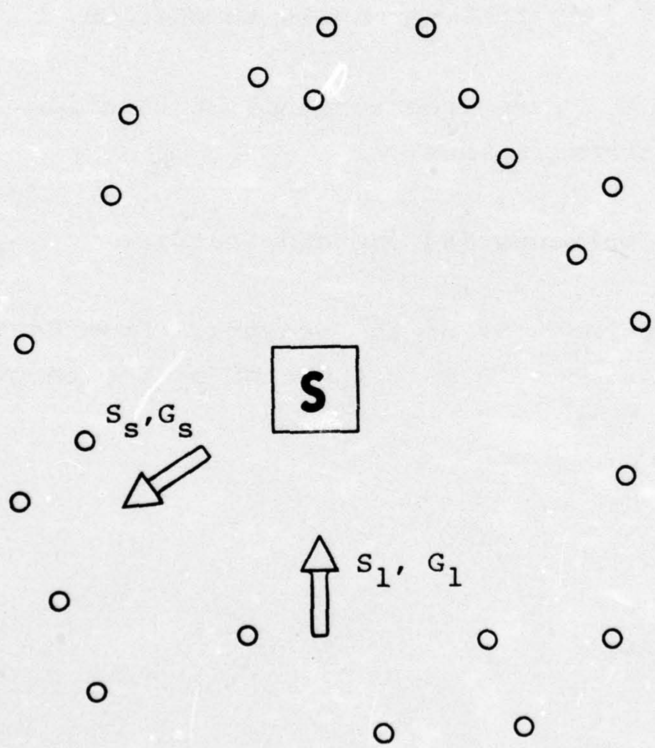


FIGURE 3.2: 1-HOP, INFINITE TERMINAL POPULATION MODEL

Where:

S_s = Effective data rate from station to terminals (packets/slot).

G_s = Transmission data rate from station to terminals (including retransmissions).

S_1 = Effective data rate from terminals to station.

G_1 = Transmission data rate from terminals to station (including retransmissions).

α = Ratio between outbound and inbound traffic.

Let $R = S_1 + S_s$ be the throughput of the network. From Equation (3.1) we can derive the expression of R as a function of the terminal transmission rate G_1:

$$R = (1+\alpha)\,\frac{G_1 e^{-G_1}}{1+\alpha G_1} \tag{3.2}$$

The maximum throughput is:

$$R^* = \max_{\text{over } G_1} R(G_1) = e^{-G_1}(1-G_1 G_s)$$

Where: $G_1^* = 1-G_s^*$

$$G_s^* = \frac{1+2\alpha-\sqrt{1+4\alpha}}{2\alpha}$$

At α=1, the maximum throughput R* = .4. The optimal transmissi~~ parameters are:

$G^*_l = .619$
$G^*_s = .381$
$S^*_l = .205$
$S^*_s = .205$

Recalling that the average number of transmissions per packet B is B = G/S, we notice that, at maximum throughput conditions, a packet from station to terminal is transmitted on the average two times, while a packet from terminal to station is transmitted three times. This is not surprising, since inbound packets are more likely to be blocked than outbound packets (see Eq. (3.1)).

In order to achieve the maximum throughput as in Eq. (3.2), station and terminals must properly "balance" the respective transmission rates G_s and G_l. If the rates are not balanced, the throughput may be considerably less than the value indicated by (3.2). As an example, consider a PR system with a terminal population transmitting with rate G_l = .61. The station must then respond with one transmission (or retransmission) every 3.2 slots in order to achieve maximum throughput. If the station is too agressive, i.e., it transmits at a rate higher than optimum (e.g., too short station retransmission intervals), then it will tend to flood the network and reduce the effective throughput from terminals to station. If, on the other hand, the station transmits at a rate lower than optimum (e.g., too long retransmission intervals) the station queue will rapidly overflow causing severe throughput degradation. In particular, if all the traffic is store and forward, packets can reach the station more easily than leave it. This leads to buffer overflow, packet discard and end-to-end retransmission with consequent throughput reduction.

To illustrate the impact of unbalanced transmission ratio's on throughput, we report below the value of R for various values of G_1 and G_s. In Table 3.1, Network Throughput R is calculated assuming $\alpha = 1$, i.e.:

$$R = 2 \times \{\min (S_1, S_2)\}$$

To interprete Case 1, let us consider a query/response traffic situation in which the queries originate in a higher level net and are directed to Hosts installed in the PRNET. Queries enter the PRNET through the station gateway, and responses are returned by the Hosts to the higher level net via the gateway. Since $S_s > S_1$, more queries are successfully received by terminals than responses returned. This may be due to excessive delays encountered by response packets on the critical hop from terminal to station, causing several queries to be timed out and retransmitted by the station before a response is heard. In our example, the number of queries that must be submitted for each response is $S_s/S_1 = 2.5$. The overall throughput degradation is $\frac{.4-.26}{.4} = 40\%$. To correct this situation, the station retransmission interval must be increased so that G_s is reduced from .6 to .39.

Case 2 corresponds to a system in which station retransmission intervals are much shorter than optimum. Assuming that the traffic is store-and-forward, the fact that $S_1 > S_s$ corresponds to frequent station buffer overflows, which require end-to-end retransmission from terminals. In this specific example $S_1/S_s = 6$. Thus 5 out of 6 inbound packets are discarded because of station buffer overflow and must be retransmitted.

From these examples, it is clear that transmission rate balancing is of critical importance for efficient PRNET operations. In the sequel, we will discuss techniques for the adaptive balancing of such rates.

	G_l	G_s	S_l	S_s	R
Case 1	.61	.6	.13	.326	.26
Case 2	.61	.1	.298	.05	.1
Case 3 (optimum)	.61	.39	.2	.2	.4

TABLE 3.1: EFFECT OF UNBALANCED TRANSMISSION PARAMETERS ON THROUGHPUT R

3.2.2 One-Hop Repeater Model

Model description:

- Single station.
- m repeaters (at one-hop from station).
- Each repeater can interfere with I neighbors.
- Slotted ALOHA channel.
- Terminal to repeater communications on separate channels, with independent transmitters (i.e., no interference between terminal and repeater transmissions).
- Infinite buffer storage (in station and repeaters).

The system can be analyzed as two separate subsystems: (1) the terminal to repeater hop; and (2) the repeater to station hop. For the terminal to repeater hop a modified version of the infinite population model discussed in Section 3.2.1 will apply [GITMAN, 1974]. Here we focus on the repeater to station hop (see Figure 3.3). The steady state equations are as follows:

$$\begin{cases} S_R = G_R \left(1-\frac{G_R}{m}\right)^{m-1} (1-G_s) \\ S_s = G_s \left(1-\frac{G_R}{m}\right)^{I+1} \\ S_s = \alpha S_R \end{cases} \tag{3.3}$$

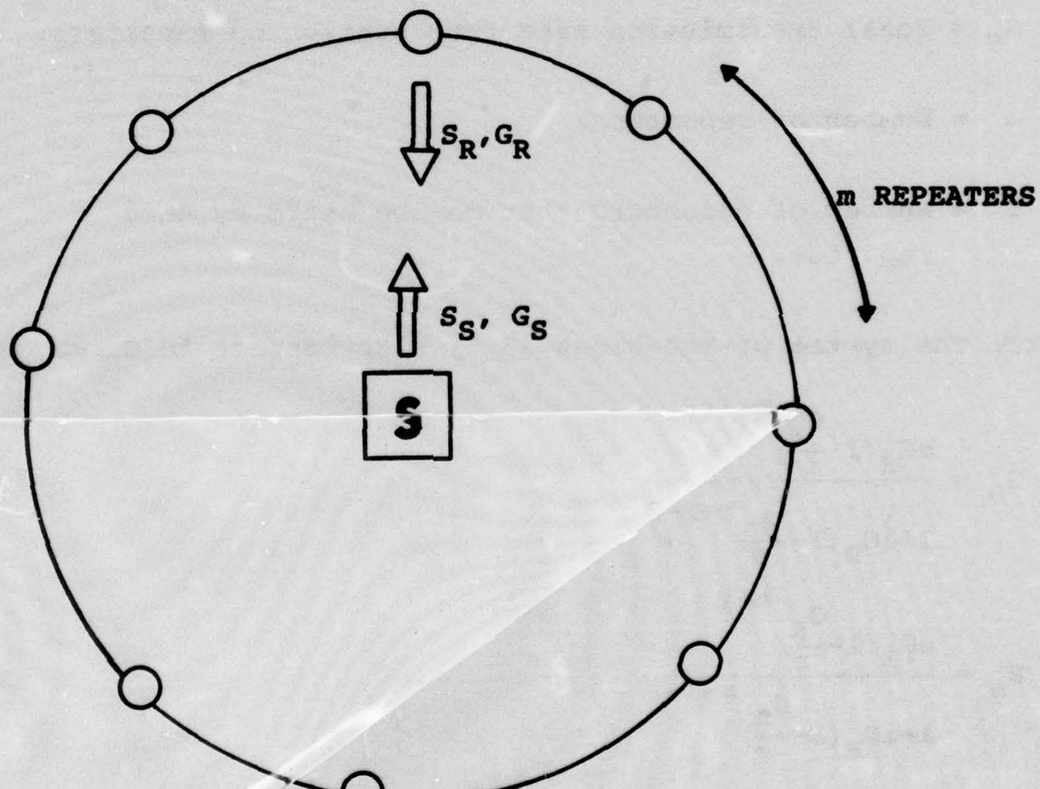


FIGURE 3.3: 1-HOP REPEATER MODEL

Where:

S_R = Total effective rate from repeaters to station.

S_S = Total effective rate from station to repeaters.

G_R = Total transmission rate from repeater to station.

G_S = Total transmission rate from station to repeaters.

m = Number of repeaters.

I = Number of neighbors that can be heard by each repeater.

Solving the system of Equations (3.3) with respect to G_R we obtain:

$$G_S = \frac{\alpha G_R \left(1-\frac{G_R}{m}\right)^{m-I-2}}{1+\alpha G_R \left(1-\frac{G_R}{m}\right)^{m-I-2}} \tag{3.4}$$

$$S_s = \frac{\alpha G_R \left(1-\frac{G_R}{m}\right)^{m-1}}{1+\alpha G_R \left(1-\frac{G_R}{m}\right)^{m-I-2}} \tag{3.5}$$

In general, we cannot obtain a closed form solution for the value G_R^* which maximizes the throughput $R = S_s(1+1/\alpha)$. However, for the special case $I = m-1$ (i.e., complete interference among all the devices), we can use the additional equation [ABRAMSON, 73]:

$$G_s^* + G_R^* = 1 \tag{3.6}$$

Equation (3.6) combined with Equation (3.4) leads to the following expression of optimal rates:

$$G_R^* = \frac{-(1+\frac{1}{m}) + \sqrt{(1+\frac{1}{m})^2 + 4\ (\alpha - \frac{1}{m})}}{2(\alpha - \frac{1}{m})}$$

$$G_S^* = \frac{2(\alpha - \frac{1}{m}) + (1+\frac{1}{m}) - \sqrt{(1+\frac{1}{m})^2 + 4\ (\alpha - \frac{1}{m})}}{2(\alpha - \frac{1}{m})}$$

$$S_S^* = G_S^* \ (1-\frac{G_R^*}{m})^m$$

It can be easily seen that the above solutions become identical to the Infinite Population Model solutions for $m \to \infty$.

Since the 1-Hop Repeater Model is essentially the finite version of the infinite terminal population model, most of the properties shown for the infinite terminal model (e.g., transmission rate balance requirement) apply here also. In addition, we may study the effect of number of repeaters m and degree of interference I on throughput performance.

In Table 3.2, we show system throughput performance $R = S_R + S_S$ as a function of number of repeaters, for $\alpha=1$ and for two different levels of interference, namely $I = m-1$ (complete interference) and $I = 2$ (nearest neighbor interference). For the complete interference model, R decreases with increasing m as expected. For $m \to \infty$ and complete interference, $R \to .411$, as in the infinite terminal population case. For $I = 2$, on the other hand,

m	R (I=2)	R (I=m-1)
3	.439	.439
4	.463	.432
5	.478	.428
6	.489	.425
7	.496	.423
8	.502	.422
9	.506	.421
10	.509	.420
∞	.538	.411

TABLE 3.2: THROUGHPUT R AS A FUNCTION OF m FOR DIFFERENT VALUES OF INTERFERENCE I

R increases with m; and for $m \to \infty$, $R \to .538$. (The limiting value for $m \to \infty$ was obtained by maximizing the limit of S_s in Equation (3.5)). From the results of Table 3.1, we may conclude that the throughput of a general 1-hop system consisting of a mix of large and small users with various degrees of interference ranges between .411 and .538, assuming that $\alpha=1$, $m \geq 3$ and $I \geq 2$.

Next, we recall that $P = G_R^* + G_s^* = 1$ for a complete interference system at maximum throughput [ABRAMSON, 1973], and wish to determine the deviation of P from unity when the interference is not complete. In order to do so, we calculate and plot in Figure 3.4, the value of P versus m for I = 2, and for various values of P is always >1 and increases with m, thus indicating that in a system with partial interference the total transmission rate at maximum throughput is higher than in a system with complete interference. Notice that P = 1 for $\alpha=0.$; and $P \to 1$ for $\alpha \to \infty$. P is maximum for $\alpha=4$. The curves in Figure 3.4 give an indication of the error introduced when applying the equation $G_R^* + G_s^* = 1$ to a system with partial interference. As it can be seen, the error is <20% for $m \leq 6$ and $.5 \leq \alpha \leq 2$ (typical parameter ranges). Thus the supplementary equation $G_R^* + G_s^* = 1$ can be used to find an approximate solution for partially interfered systems which cannot be solved otherwise.

Finally, we study system behavior as a function of α. Table 3.3 shows values of network throughput R and other transmission parameters for values of α ranging from .1 to 10. R increases with α, as expected, since the hop from station to repeaters encounters less interference than the hop from repeaters to station. B_R and B_s are the transmission ratio's defined as $B_R = G_R/S_R$ and $B_s = G_s/S_s$, and represent the average number of transmissions per packet before success in the inbound and outbound direction, respectively. Notice that B_R varies considerably for α ranging from .1 to 10.

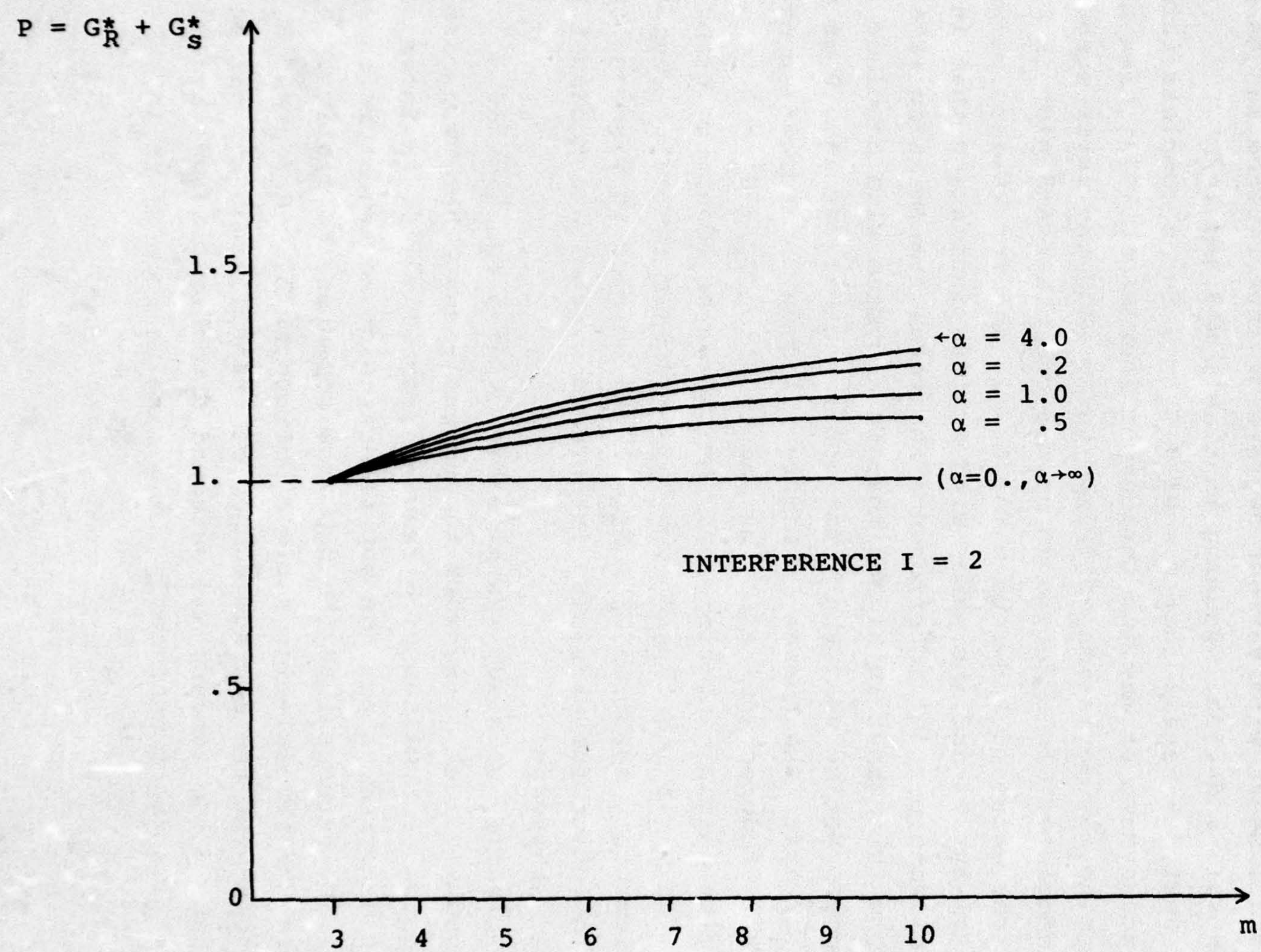


FIGURE 3.4: DEPENDENCE OF THE SUM $G_R^*+G_S^*$ ON NUMBER OF REPEATERS AND ON α, FOR I=2

α	R	G_R	G_S	B_R	B_S
.1	.413	.97	.06	2.6	1.6
.5	.452	.87	.24	2.9	1.6
1	.488	.77	.37	3.1	1.5
2	.540	.65	.50	3.6	1.4
10	.695	.37	.77	5.8	1.2

TABLE 3.3: MAXIMUM THROUGHPUT R AND OPTIMAL TRANSMISSION PARAMETERS AS A FUNCTION OF α, FOR $m = 6$, AND $I = 2$

Therefore, we must consider the dynamic adjustment of the values B_R and G_R (by adjusting the retransmission interval of the repeaters, for example) to changes of α, in order to maximize network throughput performance.

3.2.3 Two-Hop Model (Repeaters and Infinite Terminal Population)

Model description (see Figure 3.5):

- Single station.
- m repeaters (at one-hop from station).
- Infinite terminal population (at 2-hops from station).
- Dual data rate channels ($r=\frac{\text{High rate}}{\text{Low rate}} >1$). The low data rate channel is used for communication with terminals. The high data rate is used for communication between repeaters and station.
- Common repeater and terminal channel.
- Nearest neighbor repeater interference model (I=2).
- No direct terminal-station communications (terminals communicate with station only through repeaters).
- Partial interference from terminals to repeaters.
- No interference from station to terminals.

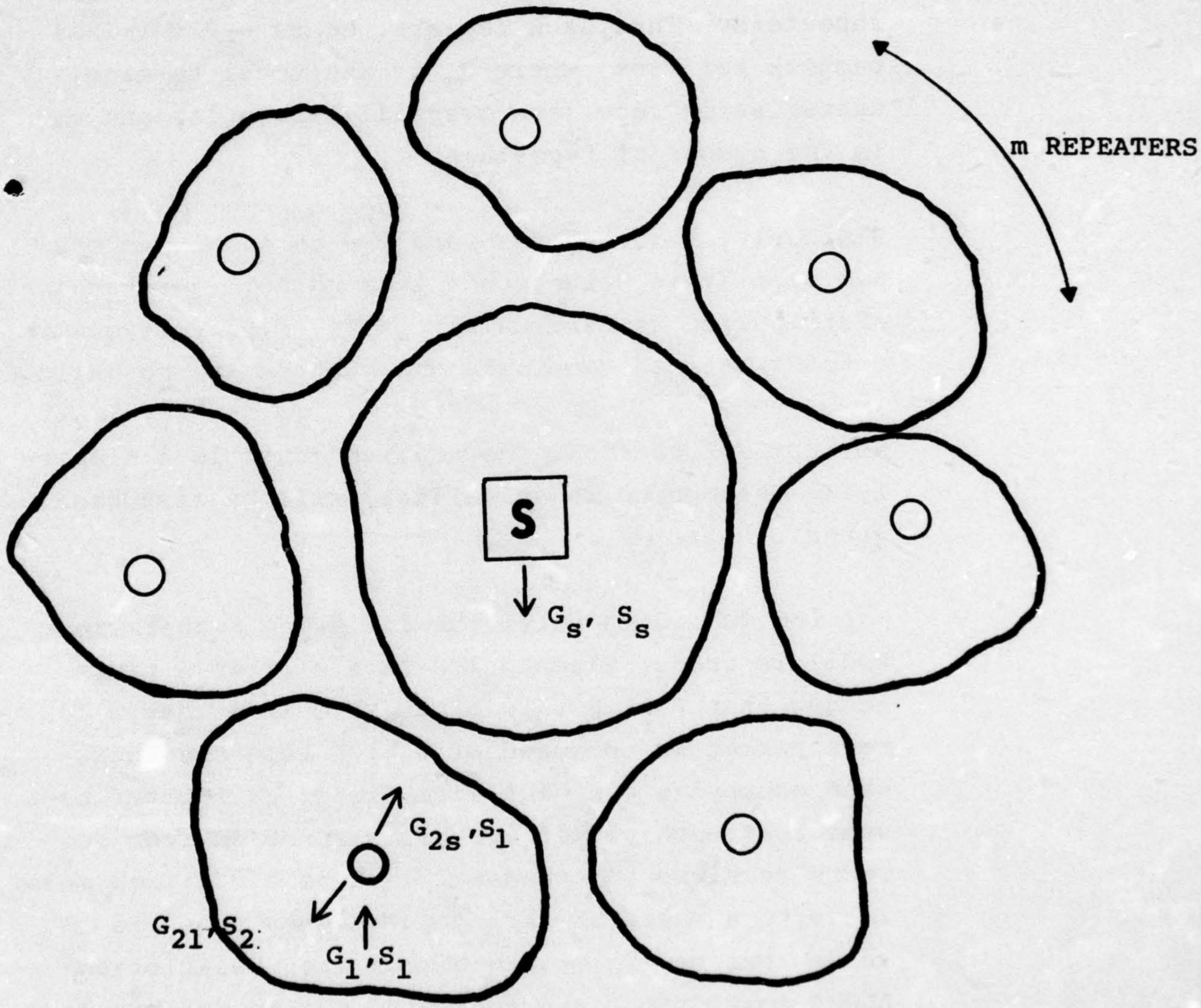


FIGURE 3.5: 2-HOP MODEL

The following assumptions are made:

- Each terminal can be heard on the average by two repeaters. Thus each repeater hears $\frac{2G_T}{m}$ terminal packets per slot, where G_T is the total terminal transmission rate (sum over all terminals) and m is the number of repeaters.

- Similarly, the station hears $\frac{2G_T}{m}$ terminal packets per slot (this corresponds to a uniform geographical distribution of terminals). However, direct communications between terminals and station are not allowed.

- No terminal can hear the station (this is a simplifying assumption whose validity will be discussed later).

- For the dual data rate, the following rather crude model is used. Since a low rate slot is 4 times longer than a high rate slot, we assume that a low rate packet is composed of r high rate segments, each occupying one high rate slot. We further assume that each packet can be reassembled from segments received independently. This simplified model leads to a value of throughput higher than the value that can be obtained from the real slotted ALOHA model since the segmented packet suffers less interference than the entire packet. However, the assumption is not too unrealistic when modeling carrier-sense systems in which low and high data rate packets suffer an equivalent level of interference.

Based on the above assumptions, the equations expressing system behavior at steady state are the following:

$$
\begin{cases}
S_1 = \dfrac{G_1 e^{-2G_1}}{r} \left(1-(G_{21}+G_{2s})\right)^3 (1-G_s) \\[2ex]
S_1 = G_{2s} e^{-2G_1} \left(1-(G_{21}+G_{2s})\right)^{m-1} (1-G_s) \\[2ex]
S_2 = \dfrac{G_s}{m} \left(1-(G_{21}+G_{2s})\right)^3 e^{-2G_1} \\[2ex]
S_2 = \dfrac{G_{21}}{r} \left(1-(G_{21}+G_{2s})\right)^2 e^{-2G_1} \\[2ex]
S_2 = \alpha S_1
\end{cases}
\tag{3.7}
$$

Where:

m = Number of repeaters.

r = Ratio of channel data rates.

G_1, S_1 = Rates from terminals to the associated repeater (Note: $G_1 = \frac{G_T}{m}$, where G_T is the total terminal rate).

G_{21}, S_2 = Rates from a repeater to the associated terminals.

G_{2s}, S_1 = Rates from each repeater to the station.

$\frac{G_s}{m}$, S_2 = Rates from station to each repeater.

The total throughput R is given therefore by:

$$R = m\, S_2\, (1+1/\alpha)$$

From the above system of equations, one cannot in general derive a closed form solution of the system variables as a function of G_1 except for m=3, 4 and 5. Let:

$$A = \frac{1+\alpha G_1(m/r+1)}{\alpha G_1}$$

and

$$C = \frac{(\alpha G_1)^{m-3}}{\alpha r(1+\alpha G_1 m/r)^{m-4}}$$

We have (for m=3,4 and 5):

$$G_{21} = \begin{cases} (A+C)^{-1} & m=3 \\ (1-C)A^{-1} & m=4 \\ \dfrac{1+\sqrt{1-4AC}}{2A} & m=5 \end{cases}$$

$$G_{2s} = \left(\frac{rG_s}{m}\right)^{m-3} \frac{1}{r\alpha G_{21}^{\,m-4}(1-G_s)}$$

$$G_s = \frac{\alpha m G_1}{r+\alpha m G_1}$$

$$S_2 = \frac{G_{21}}{r}\,(1-(G_{21}+G_{2s}))^2\, e^{-2G_1}$$

$$R = S_2\, (1+\frac{1}{\alpha})$$

A closed form expression of the maximum throughput R as a function of system parameters cannot be obtained. Therefore, numerical techniques are employed for its determination.

In Figure 3.6, we plot the throughput R as a function of the total terminal transmission rate $G_T = mG_1$ for $\alpha=1$, $r=1$, and $m \cdot 3$, 4 and 5. It may be noticed that the maximum throughput increases with m. This leads to believe that throughput is maximized for $m>5$. An upper bound on throughput performance is given by the results of the one-hop repeater model (with I=2). For such a model the maximum throughput is .53 and is obtained for $m \to \infty$. However, for large m our interference assumptions should be revised. In particular, we should consider a model with I=km $1 \geq k \geq 0$. Thus, a more realistic bound on two-hop throughput is R=.4, i.e., the value obtained with a one-hop, complete interference model.

We now study the behavior of R for values of mG_1, beyond the optimal rate. This behavior is of interest for stability considerations, as discussed in the sequel. Notice from Figure 3.6 that the throughput drops much more rapidly for m=5 than for m<5. This can be attributed to the fact that the bottleneck for m=5 is the repeater to station hop, while for m<5 the bottleneck is the terminal to repeater hop. For m<5, if terminal input rate exceeds network capacity, the overload is absorbed by the terminal to repeater hop, while the repeater to station hop does not suffer an immediate degradation. For m=5, on the other hand, the input overload produces an immediate increase of congestion on the repeater to station hop, with rapid degradation of global performance. A clear symptom of congestion on the repeater to station hop is the rapid increase of G_{2S} versus G_1 for m=5 (see Figure 3.6). For m=4, $G_{2S}=G_1$, while for m=3, $G_{2S}<G_1$.

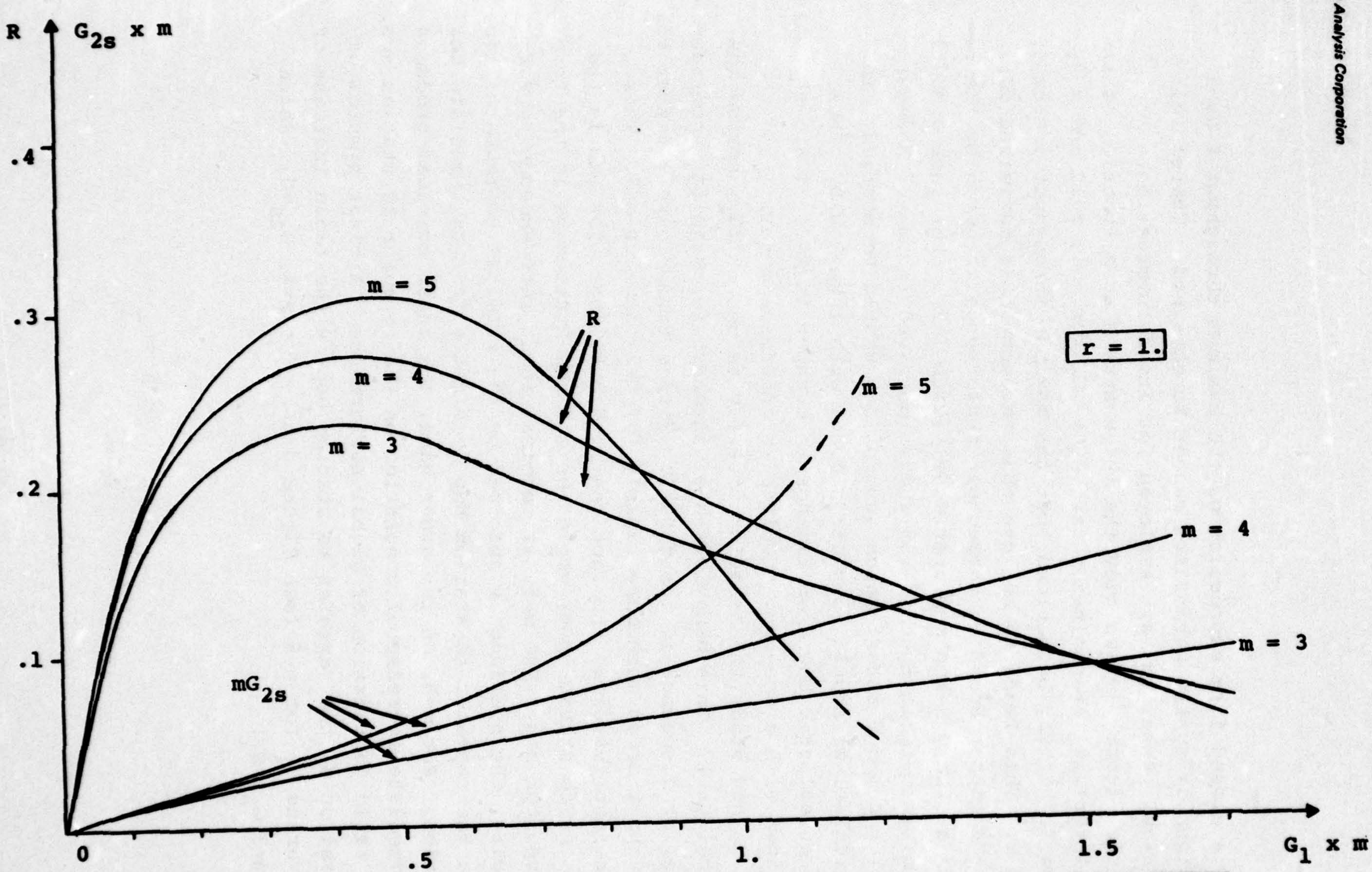


FIGURE 3.6: VALUES OF R AND G_{2s} AS A FUNCTION OF INPUT RATE G_1 x m in a 2-HOP SYSTEM

The above conclusions were derived for r=1. For r=4 we notice from Figure 3.7 that the performance degradation for m=5 is not as dramatic as for r=1. However, for $mG_1 > 3$ the throughput for m=5 decreases more rapidly than for m=3 or m=4.

Table 3.4 shows the maximum throughput R that can be achieved for different network parameters. We notice that R increases with m (as discussed before). There is a slight increase of R with α, for r=1. For values of r>1, R seems to be rather insensitive to variations of . There is a strong dependence of R on r, as expected. The throughput for r=4 is about 2/5 of the throughput that can be obtained with r=1, assuming that the high data rate is the same in both systems.*

We now introduce the notion of bottleneck (or critical hop), to identify the section of the PRNET that is most likely to become congested under heavy load. Let us first consider a multihop PRNET (e.g., a two-hop net with repeaters at one-hop from station and terminals at two-hops from station) and let us assume that the network operates at maximum throughput. Letting S_{ij} and G_{ij} be the effective rate and the transmission rate, respectively from level i to level j (e.g., from terminals to repeaters in the above-mentioned two-hop case), we recall that the average number of transmissions until success from i to j is given by $B_{ij} = G_{ij}/S_{ij}$. B_{ij} is therefore a measure of the interference encountered in forwarding a packet over the hop from level i to level j. The critical hop may be defined as the hop with maximum interference, i.e., maximum B_{ij}. From this definition it follows that the critical hop has the property of presenting the largest queues and the highest probability of buffer overflow. When the input load is increased beyond network capacity, the excess load is discarded mostly in correspondence to the critical hop. The critical hop and its position in the network (i.e., the level at which it occurs) have an important

* The above throughput results were obtained by assuming the high data rate fixed and by varying the low data rate. Therefore, an increase of (high rate)/(low rate) ratio r implies a decrease of low rate and thus a decrease in throughput. If we assume, on the other hand, that the low rate is fixed and the high rate varies, then the throughput values reported above must be multiplied by r. In the latter case the throughput increases with r.

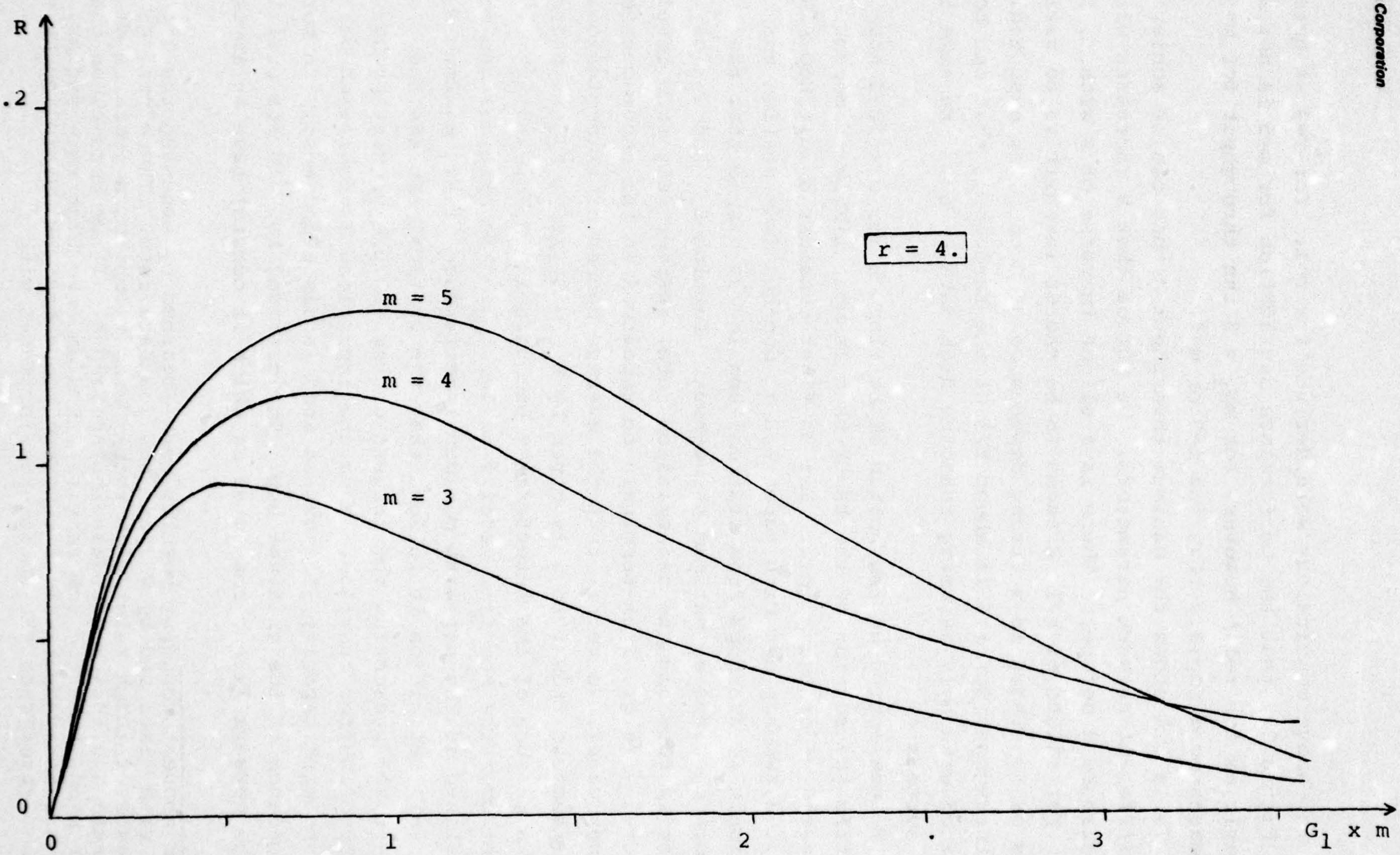


FIGURE 3.7: VALUES OF R AS A FUNCTION OF INPUT RATE G_1xm, in a 2-HOP SYSTEM WITH r = 4

		m=3	m=4	m=5
α=1	r=1	.23	.27	.30
	r=2	.16	.19	.22
	r=3	.12	.15	.17
	r=4	.09	.12	.14
α=2	r=1	.25	.29	.33
	r=2	.16	.20	.23
	r=3	.12	.15	.18
	r=4	.09	.12	.14

TABLE 3.4: MAXIMUM THROUGHPUT R AS A FUNCTION OF NETWORK PARAMETERS

impact on network stability. We have seen, for example, that when the bottleneck is the hop from repeaters to station, the tail of the R vs G curve tends to drop much faster than when the bottleneck is the terminal to repeater hop. This property will be considered in the design of stability procedures.

The values of B_{ij} were calculated for various network parameters and are reported in Table 3.5. For m=3, the bottleneck is the terminal to repeater hop. For m=4, the terminal to repeater and the repeater to station hop become critical at the same time. For m=5, the critical hop is the repeater to station hop. The outbound hops (station to repeater and repeater to terminal) never become critical, and their transmission ratios are considerably lower than the inbound ratios. This in part justifies our initial assumption of no station to terminal interference. Even accounting for such interference the repeater to terminal hop would not become the critical hop. Thus, the assumption has no major impact on network performance.

The fact that the bottleneck moves from the terminal hop to the repeater hop when the number of repeaters m is increased beyond a certain value can be explained with the following argument. If m increases, the values of S_1 and G_1 (transmission rates from terminals to each repeater) decrease, thus reducing the interference at the terminal level. The throughput (and therefore interference) at the station, on the other hand, increases with m. We may expect therefore that B_{12} decreases with m and B_{2s} increases with m. The actual results in Table 3.5 show that B_{12} decreases very slowly with m, while B_{2s} increases quite rapidly.

Additional insight into bottleneck and critical hop properties can be gained by studying the behavior of the total channel traffic P as a function of m. The notation of P was first introduced in

$\alpha=1$ $r=1$	m=3	m=4	m=5
B_{12}	3.3*	2.94*	2.92
B_{2S}	2.54	2.94*	3.47*
B_S	2.40	2.05	2.01
B_{21}	1.95	1.76	1.69
R	.23	.27	.30

$\alpha=2$ $r=1$	m=3	m=4	m=5
B_{12}	3.58*	3.26*	3.21
B_{2s}	2.90	3.26*	3.76*
B_s	2.24	1.99	1.88
B_{21}	1.83	1.65	1.6
R	.25	.29	.33

B_{12} = Number of Transmissions From Terminal to Repeater.

B_{2s} = Number of Transmissions From Repeater to Station.

B_s = Number of Transmissions From Station to Repeater.

B_{21} = Number of Transmissions From Repeater to Terminal.

R = Throughput.

* = Critical Hop.

TABLE 3.5: AVERAGE NUMBER OF TRANSMISSIONS BEFORE SUCCESS FOR A 2-HOP SYSTEM, FOR VARIOUS VALUES OF α, r AND m

$\alpha=1$ $r=4$	m=3	m=4	m=5
B_{12}	2.98*	2.92*	2.95
B_{2s}	2.45	2.92*	3.54*
B_s	2.59	2.46	2.44
B_{21}	2.15	2.04	2.03
R	.095	.12	.14

$\alpha=2$ $r=4$	m=3	m=4	m=5
B_{12}	3.24*	3.17*	3.3
B_{2s}	2.56	3.17*	4.1*
B_s	2.67	2.44	2.5
B_{21}	2.11	2.01	2.01
R	.096	.12	.145

TABLE 3.5: (CONTINUED)

Section 3.2.2 to define the average number of packet transmissions per slot measured in maximum throughput conditions. For the two-hop system the value of P depends on the hierarchical level. At the station level, P_s is given by:

$$P_s = 2G_R + m\ (G_{21}+G_{2s}) + G_s$$

As a repeater level, P_R is given by:

$$P_R = 2G_R + 3\ (G_{21}+G_{2s}) + G_s$$

Recall that P=1 for a system with complete interference.

The values of P_s and P_R for r=1, α=1 and m = 3, 4 and 5 are reported in Table 3.6. Notice that P_R decreases and P_s increases with m. This is a further indication of the fact that the bottleneck is gradually shifting from the terminal to the repeater hop. Furthermore, the deviation of P from unity is rather small, and is consistent with the results obtained in Section 3.2.2.

3.2.4 The Multihop Model

We now consider the more general case of a PRNET with a large population of terminals with uniform georgraphical distribution and a finite number of repeaters which provide area coverage and are organized in a hierarchical multi-level structure (see Figure 3.8). The exact throughput and performance analysis of a multihop system is very cumbersome, and there is little hope to get close form solutions except for very special cases. We therefore follow an approximate approach which is an extension of the two-hop model discussed in Section 3.2.3.

We first assume that the critical hop is the hop nearest to the station, so that for performance evaluation purposes we only need to consider repeaters and terminals within two-hops from the

m	P_R	P_s
3	1.086	1.086
4	.965	1.125
5	.946	1.264

TABLE 3.6: TOTAL CHANNEL TRAFFIC P_R AND P_S MEASURED AT REPEATER AND STATION RESPECTIVELY, AS A FUNCTION OF NUMBER OF REPEATERS m

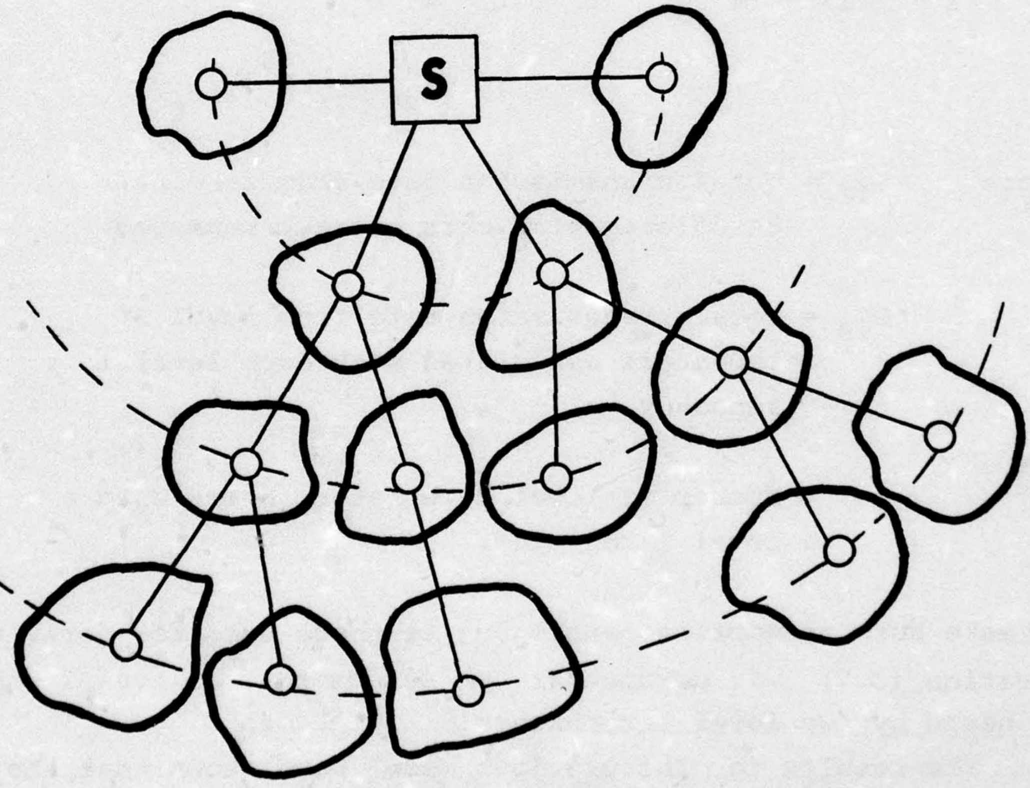


FIGURE 3.8: MULTIHOP PRNET WITH REPEATERS ORGANIZED IN A HIERARCHICAL STRUCTURE

station, (i.e., the only ones interfering with the first hop). This equivalent two-hop model differs from the model discussed in Section 3.2.3 in that the second hop contains both a large number of terminals (with low data rate) and a finite number of repeaters (with high data rate), while in the previous model the second hop consisted only of terminals.

For the equivalent two-hop model the effective rate S_1 from level 2 to level 1 can be expressed as follows (see Equation (3.7)):

$$S_1 = \left(G_{1R}\left(1-\frac{G_{1R}}{k}\right)^{-1} + \frac{G_{1T}}{r}\right)\left\{\left(1-\frac{G_{1R}}{k}\right)^{2k} e^{-2G_{1T}} \times \left(1-(G_{21}+G_{2s})\right)^{3}(1-G_s)\right\} \tag{3.8}$$

Where G_{1T} = Total transmission rate from terminals associated with each level 1 repeater.

G_{1R} = Total transmission rate from level 2 repeaters associated with each level 1 repeater.

k = Number of level 2 repeaters homed onto a level 1 repeater.

We make here assumptions analogous to those used for deriving Equation (3.7). In particular, we assume that a level 2 repeater is heard by two level 1 repeaters.

The results for the previous model have shown that the optimal value of $G_1 = G_{1R}+G_{1T}/r << 1$. Thus, recalling that for $x<<1$:

$$(1-\frac{x}{k})^{2k} \cong e^{-2x}$$

$$(1-\frac{x}{k})^{-1} \cong 1$$

We rewrite equation (3.8) as follows:

$$S_1 = \frac{G_1}{r'} e^{-2G_1} (1-(G_{21}+G_{2s}))^3 (1-G_s) \qquad (3.9)$$

Where:

$$G_1 = G_{1T} + G_{1R}$$

$$r' = \frac{r(G_{1T}+G_{1R})}{r\,G_{1R}+G_{1T}} = \frac{r(G_{1T}/G_{1R}+1)}{r + G_{1T}/G_{1R}}$$

$$= \frac{r(S_{1T}/S_{1R}+1)}{r + S_{1T}/S_{1R}}$$

Comparing Equation (3.7) with Equation (3.9) we recognize that the results derived in Section 3.2.3 apply provided that the values of G_1 and r' as defined above are used.

Let us now consider some special cases. If there is only one level of repeaters, then $S_{1R} = 0$ and r' = 4, and we have the same case studied in Section 3.2.3. If on the other hand there is a large number of repeater levels (large PRNET), then $S_{1T} \rightarrow 0$ and r' = 1. For intermediate cases we calculate r' from the ratio S_{1T}/S_{1R}, which is known from the input requirements.

In summary, approximate results for the multihop case can be obtained as an extension of the 2-hop model. In particular, the throughput versus input transmission rate behavior and the critical hop tradeoffs demonstrated for the 2-hop case can be extended to the multihop case.

3.3 STABILITY CONSIDERATIONS

3.3.1 Stable and Unstable Equilibrium

In studying the stability of a multihop system we follow the approach of Kleinrock and Lam for the single hop ALOHA broadcast channel [KLEINROCK, 1974]. The system under consideration is slotted ALOHA, and is characterized by the throughput versus input transmission rate behavior. Such behavior varies from system to system and will in general depend on geometrical network characteristics (number of hierarchical levels, number and geographical distribution of repeaters, etc.), hardware configuration (ratio of data rates, etc.), and link protocols (retransmission timeouts, etc.). A typical throughput versus input rate curve is shown in Figure 3.9.

The user population model is represented by a finite number N of active Hosts and/or terminals engaged in interactive communications. (Note: the model can be extended to include file transfers if end-to-end protocols are properly specified). Each user generates a new packet directed to a higher level network or to another user or process (e.g., a query to a data base) every T time slots. If the response from the destination is not received before a timeout of K_o slots, the user retransmits the packet after a random time interval uniformly distributed between 0 and K_1 slots. The response may not be received because a collision in the hop from terminal to repeater or because the query, although correctly received by the first repeater, was lost on its way to the station (e.g., it was discarded after maximum number of retransmissions), or because the response was lost on its way from station to terminal.

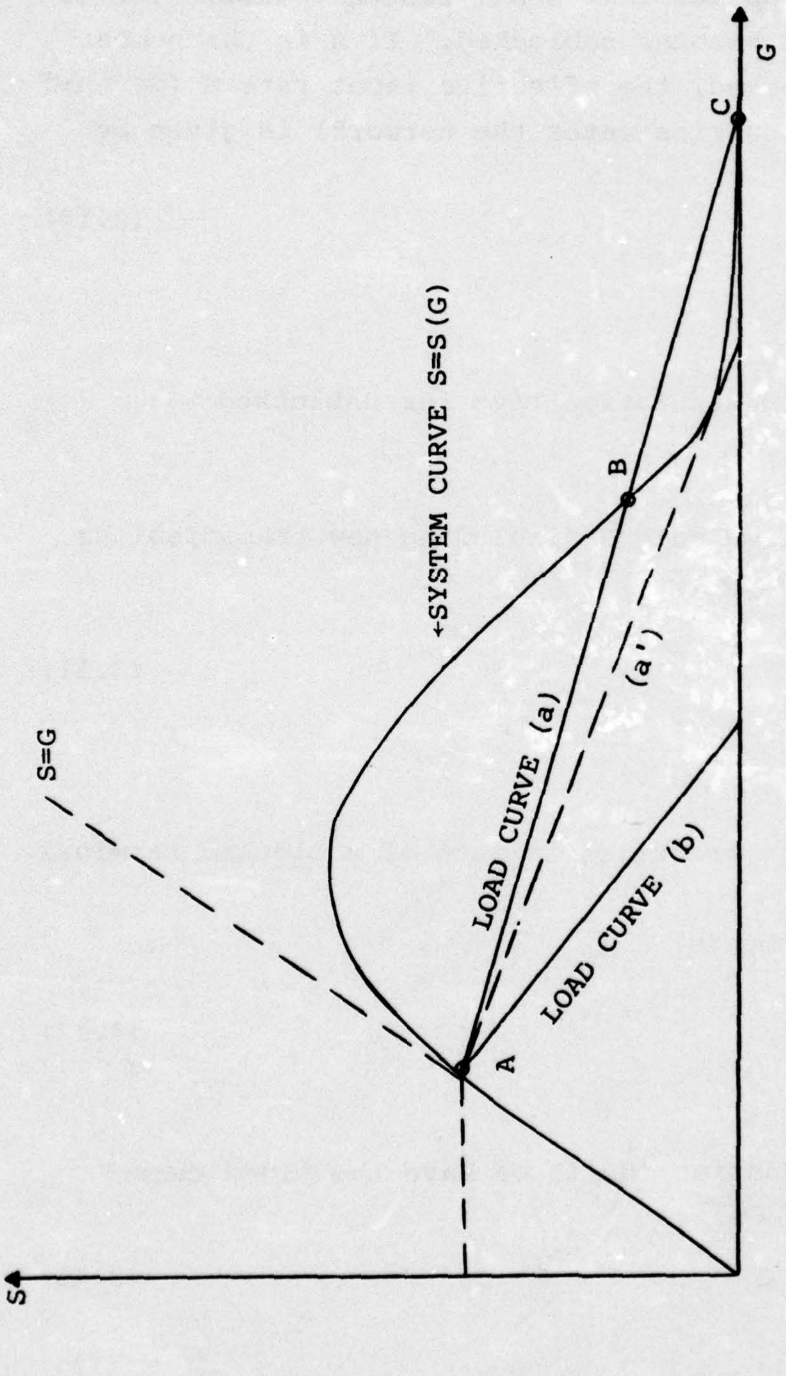


FIGURE 3.9: SYSTEM CURVE, LOAD CURVES, AND EQUILIBRIUM POINTS

A terminal engaged in retransmissions after timeout is defined to be "blocked". A blocked terminal shall resume transmission of new queries only after it becomes unblocked. If n is the number of terminals that are blocked, the effective input rate S (defined as the rate at which new queries enter the network) is given by:

$$S = (N-n)\sigma \tag{3.10}$$

Where:

$\sigma = \frac{1}{T}$ = Average query generation rate for unblocked terminals.

The total input transmission rate G (including new transmissions and retransmissions) is given by:

$$G = (N-n)\ \sigma + np \tag{3.11}$$

Where:

$p = \frac{1}{K_o+K_1/2}$ = Average transmission rate of a blocked terminal.

From Equation (3.11) we obtain:

$$n = \frac{G-N\sigma}{p-\sigma} \tag{3.12}$$

Substituting (3.12) in Equation (3.11) we have the "load curve" $S = S(G)$:

$$S = N + \frac{\sigma^2 N}{p-\sigma} - \frac{\sigma}{p-\sigma} G \tag{3.13}$$

(where $S \leq G$)

The load curve relates the effective input rate S to the input transmission rate G, and depends solely on user behavior and requirements (namely, number of active terminals, interarrival time T and retransmission timeouts). The load curve is independent of PRNET characteristics.

Our definition of load curve is slightly different from that presented in [KLEINROCK, 1974]. Effective rate S is expressed in terms of G rather than n as in [KLEINROCK, 1974], in order to relate stability considerations directly to the performance results obtained in Section 2, where G was assumed to be the independent variable. Similar results, however, may be obtained using the number of blocked terminals n to characterize the state of the system.

In Figure 3.9, we show the system curve and two load curves corresponding to different retransmission timeout values. The equilibrium points (for which effective input rate = delivered rate) are clearly the intersections of the two curves. Using the "fluid approximation" concept described in [KLEINROCK, 1974] one can easily show that A and C are the stable equilibrium points for curve (a), while B is unstable. For curve (b), A is the only equilibrium point, and is stable. In the case of curve (a), the normal operating point is A. However, input load fluctuations (e.g., a temporary increase in the number of active terminals) may drive the system to C. Indeed, this will in general occur with probability 1 if both number of active terminals and interarrival time between queries are random variables [KLEINROCK, 1974].

Various techniques have been proposed to stabilize unstable random access systems in order to maintain high performance in spite of load fluctuations [FERGUSON, 1975], [LAM, 1974], [FAYOLLE, 1974]. These techniques are generally based on the dynamic control of terminal transmission rates and/or retransmission timeouts. For

example, a technique that guarantees stability is the progressive increase of the end-to-end retransmission timeout after each retransmission [FERGUSON, 1975]. Another proposed technique consists of dynamically adjusting the timeout as a function of retransmissions and round trip delay measurements [CERF, 1974].

To illustrate the effects of the dynamic timeout adjustments consider load curve (a) in Figure 3.9 and assume that a load fluctuation has driven the system to equilibrium point C. By increasing the timeout we progressively decrease the retransmission rate of blocked terminals and move the load curve (see Equation 3.13) from configuration (a) to configuration (a'), at which point we return to A, the only point of equilibrium for load curve (a'). Once in A, the operation with the original value of timeout may be resumed.

The above control procedures presuppose the capability of adjusting (locally or remotely) the transmission parameters in each terminal. In some applications, however, terminals may not be sufficiently sophisticated to possess this capability. Or, some users may not comply with the timeout adjustment procedure in a selfish attempt to capture a larger portion of network bandwidth. Or, even worse, some users may intentionally transmit at a very high rate in order to jam the system.

These considerations suggest the need for stability procedures which are implemented in the PRNET station and repeaters, in addition to procedures implemented in the terminals. We refer to such procedures as "intranet" control procedures, as opposed to "user" control procedures, which are based on the control of user end-to-end retransmission parameters.

3.3.2 Intranet Stability Control

Intranet procedures are based on internal protocols between repeaters and station and have the purpose of improving the intrinsic system stability in addition to (but independently of) user procedures. Intranet procedures have a direct impact on the system curve $S = S(G)$ characterizing the system at steady-state. Indeed, we can derive a qualitative indication of the effectiveness of such procedures.

Figure 3.10 illustrates this concept. Let A be the original system curve and B and C be curves obtained with different control implementations. Curve B shows better stability than curve A since higher peaks of higher load fluctuation are required to drive system B to an unstable condition. Similarly, system C is more stable than system A since it permits load curves with steeper S-G slope (i.e., shorter end-to-end timeout and therefore lower delay) without becoming unstable.

Let us first investigate the control procedures that improve peak throughput performance. As discussed in Section 3.2, the throughput performance is optimized by proper selection of repeater transmission rates. The optimal selection is generally a function of input rates, ratios between inbound and outbound traffic, etc. Since input rates and traffic ratios vary in time, the adjustment must be dynamic. A possible control scheme may therefore consist of the periodic adjustment of repeater retransmission parameters from the station based on traffic and performance measurements.

Next, we investigate techniques to reduce the slope (absolute value) of $S = S(G)$, for large values of G, in order to enhance stability (see Figure 3.10). To this end, we show that the decrease of S with G for large values of G can be made less pronounced if an upperbound is imposed on the rate of packets accepted by repeaters

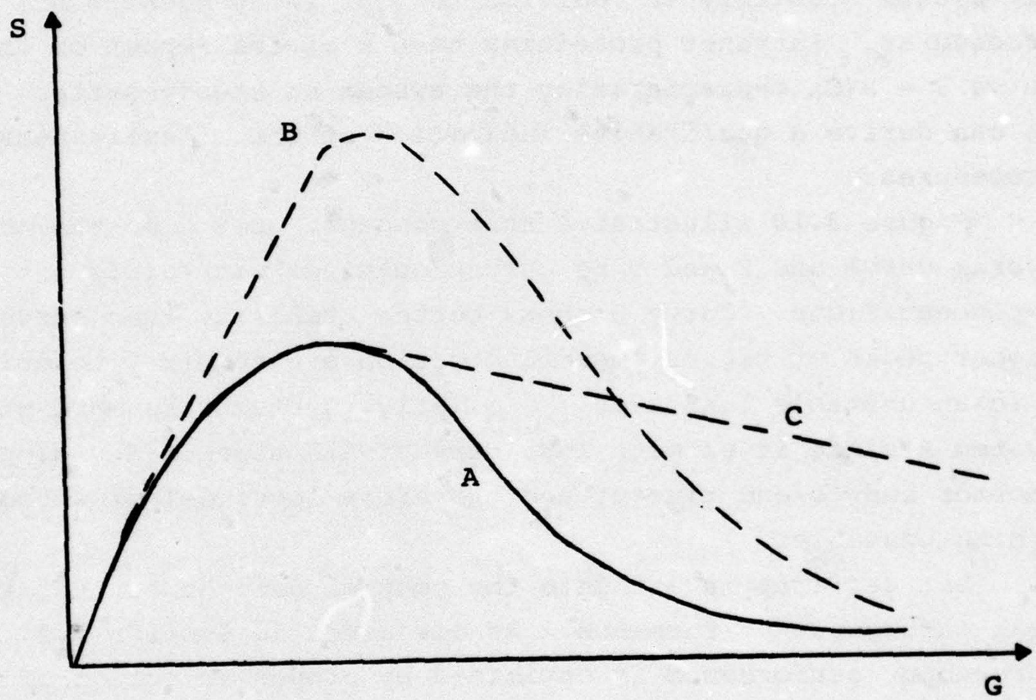


FIGURE 3.10: EFFECT OF INTRANET CONTROL PROCEDURES ON STABILITY

from the associated terminals. To prove this property we consider the two-hop system discussed in Section 3.2.3 and rewrite the equilibrium equations as in (3.7):

$$S_1 = \frac{G_1 e^{-2G_1}}{r} \left(1-(G_{21}+G_{2s})\right)^3 (1-G_s)$$

$$S_1 = G_{2s} e^{-2G_1} \left(1-(G_{21}+G_{2s})\right)^{m-1} (1-G_s)$$

$$S_2 = \frac{G_s}{m} \left(1-(G_{21}+G_{2s})\right)^3 e^{-2G_1}$$

$$S_2 = \frac{G_{21}}{r} \left(1-(G_{21}+G_{2s})\right)^2 e^{-2G_1}$$

$$S_2 = \alpha S_1$$

Let G_1^* be the terminal transmission rate that yields maximum throughput, and assume now that repeaters apply the following input rate control procedure:

1. If $G_1 \leq G_1^*$ all packets successfully received by a repeater from its terminals are forwarded to the station.

2. If $G_1 > G_1^*$, only a fraction G_1^*/G_1 of the packets successfully received is forwarded to the station, while the remaining fraction $(G_1-G_1^*)/G_1$ of successfully received packets is discarded.

Based on the above input rate regulation model we modify (3.7) to obtain the following equilibrium equations for $G_1 > G_1^*$ (note that for $G_1 \leq G_1^*$ equations (3.7) still hold):

$$S_1 = \frac{G_1^* e^{-2G_1}}{r} (1-(G_{21}+G_{2s}))^3 (1-G_s)$$

$$S_1 = G_{2s} e^{-2G_1} (1-(G_{21}+G_{2s}))^{m-1} (1-G_s) \tag{3.14}$$

$$S_2 = \frac{G_s}{m} (1-(G_{21}+G_{2s}))^3 e^{-2G_1}$$

$$S_2 = \frac{G_{21}}{r} (1-(G_{21}+G_{2s}))^2 e^{-2G_1}$$

$$S_2 = \alpha S_1$$

Letting $G_1 = G_1^* + G_x$, we rewrite (3.14) in the following form:

$$S_1 e^{2G_x} = \frac{G^*{}_1 e^{-2G_1^*}}{r} (1-(G_{21}+G_{2s}))^3 (1-G_s)$$

$$S_1 e^{2G_x} = G_{2s} e^{-2G_1^*} (1-(G_{21}+G_{2s}))^{m-1} (1-G_s) \tag{3.15}$$

$$S_2 e^{2G_x} = \frac{G_2}{m} (1-(G_{21}+G_{2s}))^3 e^{-2G_1^*}$$

$$S_2 e^{2G_x} = \frac{G_{21}}{r} (1-(G_{21}+G_{2s}))^2 e^{-2G_1^*}$$

$$S_2 = \alpha S_1$$

By comparing (3.7) with (3.15) we conclude that at equilibrium $S_1 e^{2G_x} = S_1^*$ is the inbound throughput obtained with $G_1 = G_1^*$.

Thus, for a two-hop system with input rate control we have:

$$S_1\ (G_1) = S_1^*\ e^{-2(G_1-G_1^*)} \quad \text{for } G_1 > G_1^*. \tag{3.16}$$

Using Equation (3.16) we plot in Figure 3.11 the throughput versus input rate curve for a two-hop *input rate controlled* system with r = 1, α = 1 and m = 5, and compare it with the original curve without rate control previously reported in Figure 3.6. The input controlled system displays better stability characteristics than the original system. We expect this to be true in general for all multihop systems with bottleneck localized in the hop from first level repeaters to station.

Intuitively, the above behavior is justified by observing that the same relative increase in offered rate produces lower interference and performance degradation in a lightly loaded channel (such as the terminal to repeater hop in our case) than in a saturated channel (such as the hop from repeaters to station). Therefore, we must block the increase in offered rate at a hierarchical level where the channel load is light, and prevent it from reaching the bottleneck.

To provide a quantitative evaluation of the impact of repeater input controls on stability in the above mentioned two-hop system we consider a specific numerical example. Let:

N = Number of simultaneous active terminals = 200

T = Interarrival time for new packets = 2000 slots

K_o= Timeout = 80 slots

K_1= Random retransmission interval = 40 slots

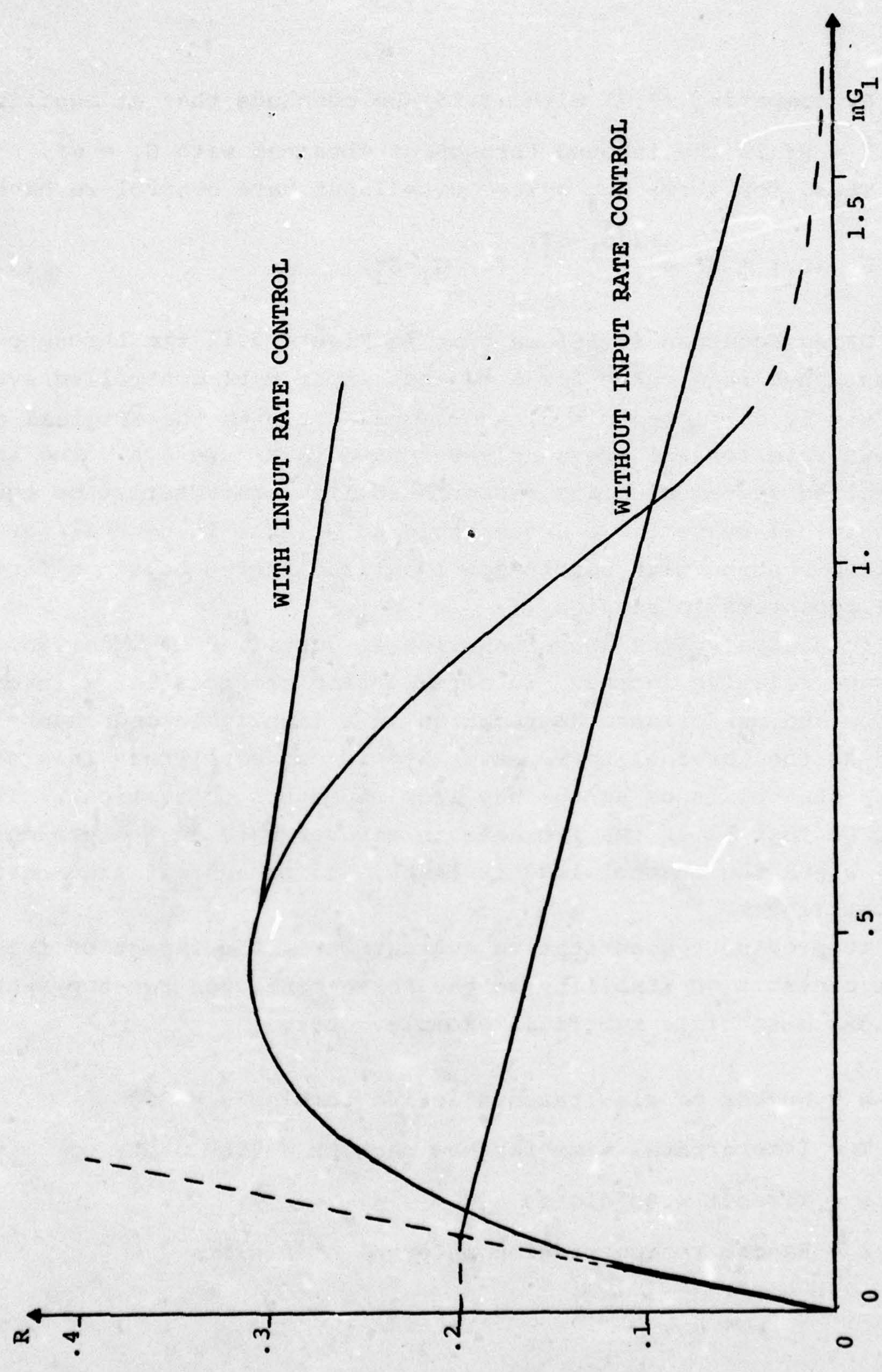


FIGURE 3.11: THROUGHPUT VERSUS TRANSMISSION RATE CHARACTERISTICS WITH AND WIHTOUT REPEATER INPUT RATE CONTROL

The load curve corresponding to the above requirements is calculated according to Equation (3.13) and is shown in Figure 3.10. We notice that the input controlled system is always stable, while the uncontrolled system is bistable and requires additional end-to-end control procedures to allow recovery from the "degraded" equilibrium condition.

In this section we have shown some basic approaches to enhance network stability using intranet controls. In the next section we describe specific procedures for the implementation of these concepts.

3.4 STABILITY CONTROL PROCEDURES

3.4.1 Function

The function of the Stability Control Procedures is to prevent the PRNET from becoming congested due to excessive traffic input rates from PR terminals or from gateways. Stability and congestion protection is obtained with the regulation of input rates and network parameters, and with the appropriate allocation (and possibly, reservation) of buffer resources in the station.

3.4.2 Characterization

PRNET congestion protection is obtained with a set of separate procedures whose operational parameters are controlled by a station stability program. A tentative classification of the procedures in functional levels is the following:

Level 1: Single-hop control procedures. The purpose is to avoid blocking of packets already accepted into the PRNET, and to make efficient use of the network resources.

Level 2: Network input control procedures. The purpose is to regulate packet inputs from terminals and gateways into the PRNET.

Level 3: End-to-end (process-to-process) control procedures. The purpose is to regulate input rates at the individual user (process) level, by controlling the operational parameters of the connection (window size, time out, etc.).

Level 4: Stability control procedure. The objective is to coordinate and control the operational parameters of the lower level procedures, and to supervise the buffer allocation in the station.

Level 1 and level 2 procedures form the category of "intranet" control procedures. Level 3 procedures may also be referred to as "user" procedures. Level 4 procedures interact with both intranet and user procedures.

In the sequel we focus our attention on intranet stability control, and therefore limit our discussion to level 1 and level 2 procedures and their interaction with level 4.

3.4.3 Single-Hop Procedures

The results of Section 3.2 show that repeater and station transmission and retransmission rates must be properly adjusted for each hierarchical hop in order to obtain optimal throughput and delay performance. Similar results are shown in Chapter 2,

where it was shown that significant performance improvements can be obtained by properly selecting the maximum number of retransmissions as the function of hop distance from station. This implies the necessity of controls on hop transmission parameters to achieve optimal delay and throughput and, consequently, stability performance.

Control may be exercised on the following repeater parameters:

1. Maximum number of retransmissions before packet discard.

2. Average time between retransmissions.

3. Maximum number of packets allowed on inbound and/or outbound output queue (if multiple buffers are available).

The control can be either distributed, or centralized, with the station acting as the central controller. In a distributed implementation each repeater adjusts its own retransmission interval K and maximum number of retransmissions P based on channel load measurements. A very simple distributed scheme may consist of progressively increasing the retransmission interval K(p) for a given packet as a function of number of previous retransmission attempts p, i.e.,

$$K(p) = (1+p)K_1$$

Where:

p = Number of previous transmissions.

K_1 = Basic retransmission interval (slots).

Since the number of retransmissions is directly related to channel congestion, the effect of the procedure is to reduce the repeater transmission rate when the channel becomes overloaded. In fact, if P is the maximum number of retransmissions allowed, the average repeater transmission rate in a congested situation (i.e., most of the packets discarded after P retransmissions) is $\frac{2}{(P+1)K_1}$ for a system with hop control; it is $\frac{1}{K_1}$ for a system without hop control.

In a centralized implementation the station dynamically adjusts the parameters K and P in all repeaters based on its own traffic and delay measurements. A possible centralized scheme may consist of increasing K and decreasing P in the first hop repeaters as a function of store-and-forward queue length in the station, with the purpose of minimizing buffer overflow probability in the station and optimizing the balance between repeater and station transmission rates over the first hop (see Section 3.2.2). A more sophisticated approach may consist of dynamically adjusting the K and P parameters as a function of hop distance from the station and of station measurements.

Distributed and centralized approaches have different advantages and disadvantages. The distributed scheme has a more prompt intervention, on a per packet basis, regardless of distance from station; while the effect of the centralized scheme is less immediate and diminishes with hop distance from station. On the other hand, the centralized scheme may achieve a better performance at steady state since it is based on a more comprehensive set of measurements available in the station.

One may also construct a hybrid scheme with both distributed and centralized features, to combine the advantages and eliminate the drawbacks of either scheme. Clearly, extensive experimentation via

simulation and field measurements is required to evaluate the merits of the various approaches and to perform fine tuning of the chosen approach.

3.4.4 Network Input Control Procedures

These procedures control the input of packets from terminals to homing repeaters (on the low data rate channel), or from the gateway to the PR communication interface in the station.

The following procedures can be considered to control terminal input to homing repeaters:

1. A station command to turn ON or OFF the low data rate channel in a repeater, to enable/disable access from (and delivery to) all terminals regardless whether connected to the station or not.

2. A station command to disable/enable in a repeater the Response to Search from terminal PR's. If the command is activated, previously assigned terminals still home onto this repeater, but new terminals cannot access it.

3. Repeater Rate Control Procedure. Each repeater is allowed to accept only up to a specified number of packets from terminals for forwarding to the station in a given time interval. A counter is set at the beginning of the interval to a value corresponding to the initial packet allocation, and is decremented by one for each packet forwarded to the station. If the counter goes to zero, i.e., the initial packet allocation is used up before the end of the interval no more packets are

accepted. However, the station may allow access to additional terminal packets (if network load permits) by promptly returning to the repeater a control packet resetting the allocation count to a value greater than zero. At the end of the interval, the allocation counter is reset to the initial value. Allocation and time interval parameters are under control of the Station Stability program.

Procedure (1) (low data rate ON/OFF) is a rather drastic form of input control and is conceived more as a measure for emergency situations than a regulation for use during normal network operations. A possible application would be that of keeping the PRNET (or sections of it) free from data packets during reinitialization of recovery from failure.

Procedure (2) (enable/disable response to search from unlabeled terminals) provides a more gradual control than Procedure (1), and is intended to prevent bursts of new terminals coming into an already heavily loaded network. Access priorities may be added to the procedure to provide selective blocking of terminal inputs.

While both Procedures (1) and (2) must be activated by station commands (and thus may fail when the station cannot deliver control packets to repeaters because of a sudden congestion), Procedure(3) is fail-safe. If the system becomes congested, no allocates are returned to repeaters and network input rates are maintained at a safe level. The procedure allocates bandwidth to repeaters dynamically according to their requirements and to the availability of network resources. Under this respect, the procedure is similar to the ARPANET source IMP/destination IMP message numbering protocol, which allows only a fixed number of messages to be outstanding in the network. Subsequent messages are blocked until end-to-end ACK's

from previous messages are received. If the network becomes congested and the resources become scarce, the end-to-end ACK's are delayed and consequently the input rates are reduced.

As a difference from the ARPANET message numbering procedure, the PRNET control procedure is much simpler, since it provides only for stability control, while sequencing and end-to-end accountability tasks are performed by higher level protocols.

The question still remains if (or when) it is more cost-effective to implement the stability control functions in the user protocols (TCP), (in conjunction with accountability and sequencing functions), or in the repeater-to-station protocols. The answer will certainly vary from application to application. Simulation experiments are being planned for a few simple cases.

3.5 CONCLUSIONS

Previous studies were addressed to the stability of the packet broadcast channel in one-hop, one-way traffic systems (e.g., ALOHA, satellite, etc.). The results obtained for such systems, however, do not completely characterize the stability behavior of more complex systems such as distributed packet radio networks. In fact, they do not offer insight into the relationship between internal network parameters and network stability.

In this study, we have focused on the dependence of network performance and stability on internal network parameters such as repeater retransmission timeouts and repeater acceptance rates, in a multihop packet radio system with traffic in both terminal-to-station and station-to-terminal direction. It was shown that network stability is directly related to the shape of the throughput versus terminal input rate curve. Approximate analytical models were developed in order to study throughput performance for a variety of network configurations:

The following important properties were demonstrated:

- In a two-way traffic system, the inbound and outbound transmission parameters must be properly "balanced" in order to achieve optimal channel utilization and thus enhance stability. In particular, retransmission timeouts of station and first level repeaters are the most critical and must be adjusted based on the ratio between outbound and inbound traffic.

- In a multihop system, the bottleneck generally corresponds to the hop from first level repeaters to the station. Network stability in heavy load conditions is greatly enhanced by regulating input rates at the terminal-to-repeater hop level.

Based on the above properties, some stability procedures for the control of hop transmission parameters and input acceptance rates were proposed.

The area of packet radio stability and, more generally, performance evaluation offers many directions for further research. Among the most critical issues we mention:

- Development and evaluation of more accurate packet radio models, including features such as carrier sense, finite buffer storage with repeaters, etc.

- Development of appropriate simulation models to validate the analytical models.

- Development of models for end-to-end protocols and flow control.

- Study of the interaction between end-to-end protocols and intranet procedures and their impact on network performance and stability.

- Stability in multistation systems.

- Stability in a system with mobile repeaters and/or terminals.

VOLUME 2

Chapter 3

REFERENCES

REFERENCES

[LAM, 1974] Lam, S., "Packet Switching in a Multi-Access Broadcast Channel With Application to Satellite Communication in a Computer Network," Report #7429, School of Engineering and Applied Science, University of California, Los Angeles, April, 1974.

[FERGUSON, 1975] Ferguson, M., "On the Control, Stability and Waiting Time in a Slotted ALOHA Random Access System," ICC 1975 Proceedings, San Francisco, July, 1975.

[TOBAGI, 1974] Tobagi, F., "Random Access Techniques for Data Transmission Over Packet Switched Radio Networks," Report #7499, University of California, Los Angeles, December, 1974.

[GITMAN, 1974] Gitman, I., R. Van Slyke and H. Frank, "On Splitting Random Accessed Broadcast Communication Channels," Seventh Hawaii International Conference on System Sciences, University of Hawaii, January, 1974.

[GITMAN, 1975] Gitman, I., "On the Capacity of Slotted ALOHA Networks and Some Design Problems," IEEE Transactions on Communications, Vol. COM-23, No. 3, March, 1975.

[ABRAMSON, 1973] Abramson, N., "Packet Switching With Satellites," NCC 1973 Proceedings, New York, June, 1973.

REFERENCES (Cont'd.)

[KLEINROCK, 1974] Kleinrock, L., and S. Lam, "On Stability of Packet Switching in a Random Multi-Access Broadcast Channel," Seventh Hawaii International Conference on System Sciences, University of Hawaii, Honolulu, January, 1974.

[FAYOLLE, 1974] Fayolle, G., et. al., "The Stability Problem of Broadcast Packet Switching Computer Networks," Proceedings of The International Workshop on Computer Architecture and Networks - Modelling and Evaluation, IRIA, August, 1974, pp. 135-140.

[CERF, 1974] Cerf, V., and R. Kahn, "A Protocol for Packet Network Intercommunication," IEEE Transactions on Communications, COM-22, No. 5, May, 1974.

VOLUME 2

Chapter 4

RELIABILITY CONSIDERATIONS IN PACKET RADIO NETWORKS

4. RELIABILITY CONSIDERATIONS IN PACKET RADIO NETWORKS

4.1 INTRODUCTION

The reliability of packet switching broadcast radio networks is addressed. The network model is that of the Packet Radio Network (PRNET) [FRANK, 1975], [KAHN, 1975], and consists of switching nodes which share a single radio communication channel. The main features which distinguish the PRNET from point-to-point packet switching networks are: (i) the communication channel is shared dynamically using a random transmission scheme, and (ii) devices use a broadcast mode of transmission so that packets can be transmitted to several devices simultaneously, and/or several packets can be simultaneously received by a receiver because of independent transmissions of several devices.

There are three types of nodes in the packet radio network: a terminal, a repeater, and a station.

In some applications the Packet Radio Station will be the interface component between the broadcast system and a point-to-point network. In addition, it will perform accounting, buffering, directory, and routing functions for the overall system. In these applications most of the traffic requirements would be from terminals-to-stations and from stations-to-terminals.

The basic function of the Packet Radio Repeater is to provide a network for connection of terminals to one or more stations, thereby increasing the size of the area that can be served by a station and providing paths to alternate stations to insure reliable communications.

Recent studies [NAC, 1974] have shown that PRNET's can easily be saturated and become unstable, unless efficient routing and flow control algorithms are used. Routing algorithms which enable point-to-point packet transportation in the broadcast network have been

proposed [GITMAN, 1976]. Given a set of repeaters which receive an identical packet, the underlying idea of the routing algorithms is to enable only one or a few repeaters which are closer to the destination to accept the packet for relaying, and have all other repeaters discard the packet.

It is apparent that a routing algorithm becomes more efficient when the number of repeaters which accept the packet for relaying is reduced, in particular in a broadcast radio network. On the other hand, from the reliability viewpoint it is desirable to have many paths to the destination which in turn suggests to increase the number of repeaters used for relaying.

The reliability of broadcast radio networks is studied in the chapter for three routing algorithms. Other design possibilities to improve network reliability are also evaluated. In particular, the location of the station, the number of stations, and the increase in repeater transmission power for obtaining higher connectivity are evaluated.

4.2 RELIABILITY MODEL

We wish to evaluate the reliability of different packet radio routing and topological design policies currently under consideration. Since the packet radio networks use a broadcast mode, it is likely that naive network routing and network design could lead to unsatisfactory reliability levels. Since network components may be placed in remote areas where portable power supplies must be used, it is likely that component failure probabilities will be higher than those in point-to-point networks such as ARPANET and failures that occur will be more difficult to correct. Therefore, reliability should be a primary design criterion.

Radio networks have certain features which distinguish their reliability analysis from the analysis of point-to-point networks. First, since they use a broadcast mode, the only physical network components are nodes. There are no physical links in the network. However, links can be thought of as existing between any two nodes between which direct communication is possible. Nodes can be of three types: stations, repeaters and terminals. Stations have a special role because nearly all communications are directed to or from them. Repeaters are vital at intermediate areas in terminal-to-station, terminal-to-terminal and possibly station-to-station communications. The large number of terminals and the possibility of their movements makes it difficult to handle them as standard network nodes.

Failures in the system are produced by a variety of mechanisms. Some types of failures have traditional properties that can be handled by standard techniques, while others may require the development of new analytical techniques. At the moment, we can foresee three types of mechanisms which may frequently generate failures. First, a large natural disturbance, such as a thunderstorm, could reduce or eliminate communications among many stations, repeaters

and terminals. Second, a temporary local power malfunction or shortage could reduce the effective range of a station, repeater or terminal. Third, a local software, hardware or power failure or a local disturbance could temporarily eliminate the communications capability of one station, repeater or terminals. The first type of failure implies that several nodes and/or links cannot be used for communications. The second implies that several links adjacent to one node cannot be used for communications. The third implies that one node cannot be used for communications. The first and second types of failures are distinctive in that the state of a given node or link is not independent of the state of another node or link. No known techniques are available for network reliability analysis without the independence assumption; however, it appears that simulation techniques could be developed for these problems. For the third type of failure it can be assumed that nodes fail independently and that links are perfectly reliable. Furthermore, this type of failure appears to be the most common and likely to be of longest duration. Consequently, we shall consider only such failures in this chapter.

Because of their large number and transitory nature, terminals will not be considered explicitly in the network. The nodes will consist of several repeaters and a station. The possibility of two or more stations will be considered; however it will always be assumed that stations are reliably connected. This would be the case if all stations were part of the ARPANET. The abstract model, then, consists of a set of nodes and links. One node is distinguished as the station node. Links are perfectly reliable. It is assumed that failure of each node is independent of the failure of any other node.

Three types of routing strategies are currently being considered for packet radio networks. Each must be modeled with a different network structure. The links in each structure are chosen from a set of possible links determined by the distance between each pair

of nodes. That is, communications (and thus a link) can exist between any two nodes less than a certain distance apart. This maximum distance is determined by the operating power of the nodes. When all nodes are operative, communication will be routed through a specified spanning tree in the network. Usually this spanning tree will be a shortest path tree rooted at the station node. Alternate routing strategies will be necessary when one or more nodes have failed.

The first routing strategy considered is tree routing. That is, we will assume that no alternate strategy is available. Although such a plan would never be used, the reliability of such a configuration is of interest in estimating the importance of using an alternate strategy.

We shall refer to the second strategy as restricted routing. Under this strategy each node in the network is assigned to a level. This level corresponds to the number of links between that node and the station node in the specified spanning tree. Communication between a particular node and the station must proceed through nodes on lower numbered levels. For example, communication from a node on level three to the station would have to proceed to a node on level two and then a node on level one and finally to the station. This routing strategy can be modeled on a directed acyclic network. All links in the network are directed from the node to another node on the next lower numbered level. In order for a pair of repeaters to communicate they both must communicate with the station. Consequently, repeater-to-repeater communication can be modeled on the same network by assuming that two repeaters talk to each other via the station.

The third strategy considered is totally adaptive routing. In this strategy communication between two nodes may proceed along any available path. This strategy can be modeled as an undirected network in which links exist between any pair of nodes less than the maximum distance apart. Figures 4.1 through 4.4 demonstrate the network structures implied by the routing algorithms.

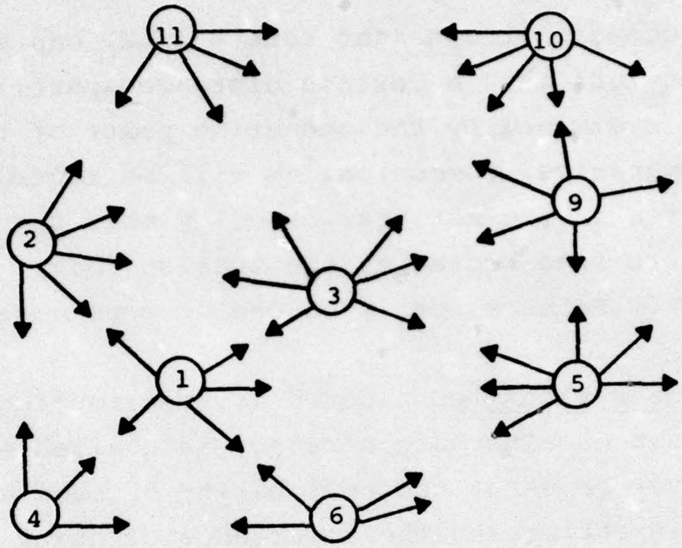


FIGURE 4.1: PACKET RADIO NETWORK: THE ARROWS FROM A NODE POINT TO NODES WITH WHICH IT CAN COMMUNICATE (NODE 1 IS CONSIDERED THE STATION)

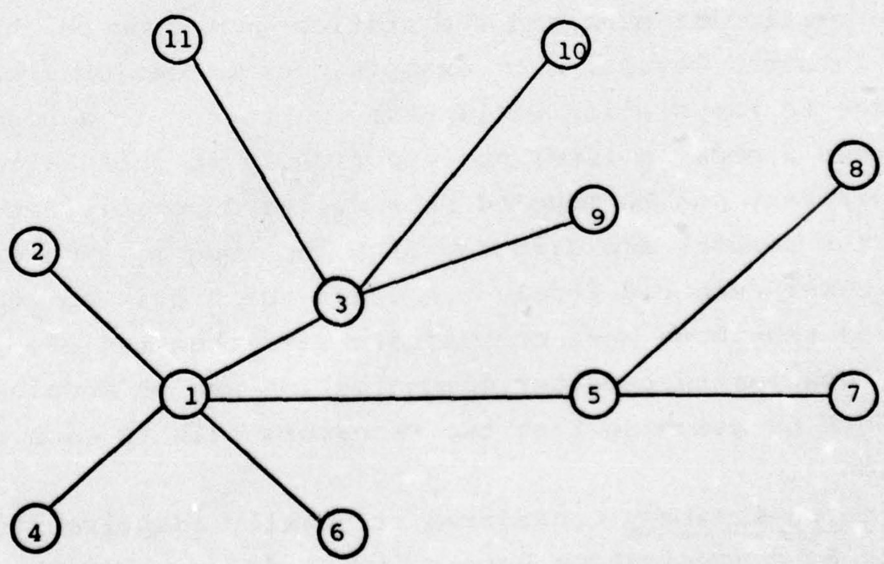


FIGURE 4.2: TREE ROUTING NETWORK

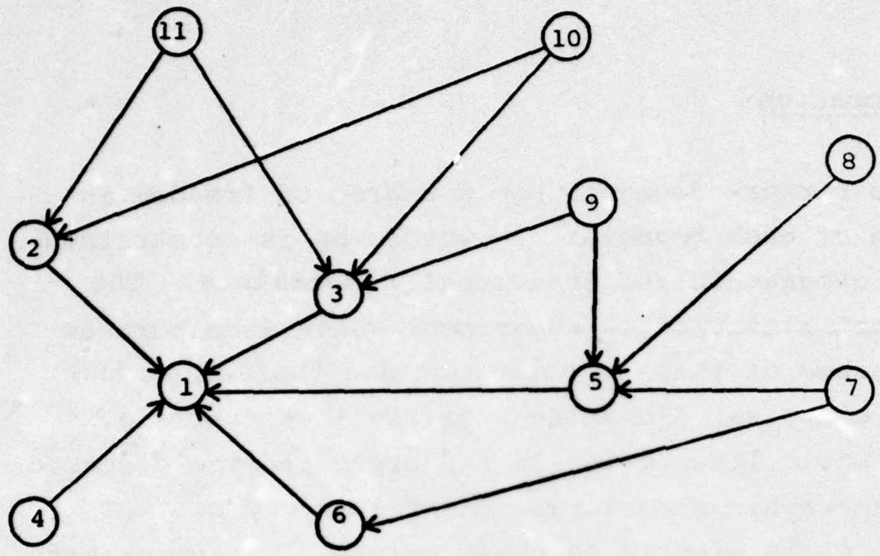


FIGURE 4.3: RESTRICTED ROUTING NETWORK

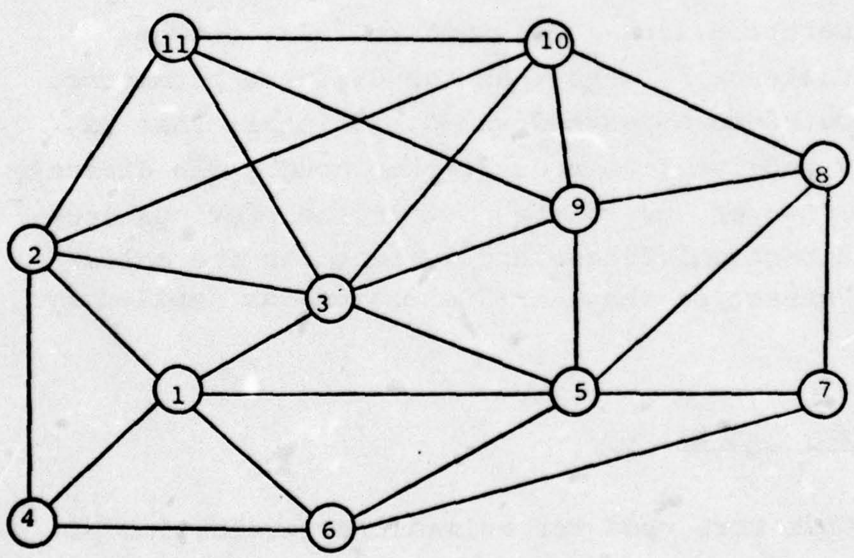


FIGURE 4.4: ADAPTIVE ROUTING NETWORK

4.3 THE ALGORITHMS

4.3.1 Network Generation

The packet radio network designer has a degree of freedom in choosing the location of each repeater. However, he is constrained to a large extent by geographic and structural constraints. The objective is to conduct a simulation experiment which incorporates both the random influence of these constraints and the choice the network designer can exercise. Assuming a fairly level terrain, the network designer would like to locate repeaters at evenly spaced points on a grid. Geographic structural constraints may prevent him from locating repeaters exactly on these points. To approximate this situation in the reliability simulator, we choose evenly spaced points on a grid and then locate repeaters randomly in rectangles drawn around such points.

The connectivity of the repeater network was determined by using a distance parameter; a link is assumed to exist between nodes whose pairwise distance is less than the distance parameter. Hence, the links generated were assumed two-directional; that is, if node i can directly receive from node j, then node j can directly receive from node i. However, as pointed out in the previous section, the links (and direction) taken into account for the reliability analysis are a subset of the links generated, as implied by the routing algorithm.

4.3.2 The Reliability Algorithms

Different algorithms were used for reliability evaluation of the different network structures implied by the routing. All the algorithms are exact. The approach taken is to enumerate a set of probabilistic events which are mutually exclusive and collectively exhaustive with respect to the reliability measure used [BALL, 1975].

The algorithms are based on a probabilistic backtracking technique which has been extensively used for solving a variety of enumeration problems. For example, suppose we propose an enumeration problem as that of enumerating all subsets of a set with a desired property. We examine elements in a prescribed order. For an element, we decide whether or not to include it in the subset under construction. When the subset has the desired property we list it. Afterwards, we change our decision about the last element and begin adding new elements until the subset again has the desired property. If changing our decision on an element cannot produce a subset with the desired property, we backup to the previous element. If this element has been considered both "in" and "out" we backup again. If it has only been considered in one state, we change our decision about it and proceed as before. When the process terminates, all subsets have been enumerated. This technique has been found very useful for determining the probability of a random occurrence, and is used in the algorithms.

The restricted algorithm [GITMAN, 1975] presented a new problem, since it implies that the network is directed and acyclic. This stimulated the development of an algorithm based on the following simple observation. The ability of a node on level L (L hops from the station) to communicate with the root (station) can be completely determined by the ability of at least one of the nodes adjacent to it on level L-1 to communicate with the root. Therefore, reliability information for level L-1 is sufficient to determine the reliability of level L. The algorithm analyzes the reliability of each level in increasing numerical order. To compute the reliability of each level, it uses only information from the previous level. At each iteration, the algorithm must convert the information concerning a communication subset on level L into information concerning a communication subset on level L+1. To perform this

analysis, the algorithm must consider all combinations of node states for the subset on level L. Each combination will yield a communicating subset on level L+1. Given the node state in the subset on level L and the subset probability and reliability measure values, the probability and reliability measure values for the communicating subset on level L+1 can be updated. In the algorithm, a probabilistic backtracking procedure is used to perform the probabilistic updates. In the context, the amount of work performed is reduced since all node state combinations need not be considered explicitly.

A tree network was used to model the tree routing. The running time of this algorithm is linear with the number of nodes. The algorithm developed for the directed acyclic network has a running time which grows exponentially with the number of nodes on a hierarchy level and linearly with the number of levels. An undirected network was used to model adaptive routing. For this model we use a general reliability algorithm with some special features incorporated for networks in which only nodes can fail. The running time of this algorithm grows exponentially with the number of nodes in the network. A detailed description of the algorithms is given in [NAC, 1976].

4.4 RESULTS OF SIMULATION STUDY

4.4.1 Introduction

A simulation was conducted using parameters which produced networks which, on the average, had 15 nodes and 30 links. These networks are smaller than most that would be encountered in practice; however, we believe that results concerning them adequately indicate general trends that can be applied to larger networks. We performed simulation studies to evaluate the effects of variations in routing strategy, variation in the placement of stations and variations in the number of stations. Studies were conducted on a few individual networks to evaluate change in reliability as a function of repeater power level and to observe the tradeoff between direct repeater-to-repeater communication as opposed to communication only through the station. Finally, in all studies we assumed that failure probability of individual nodes was constant. We conducted each study with 5 failure probability levels which ranged from .02 to .1. To evaluate the deviations when not all probabilities are equal we made a few runs in which probabilities varied from node to node.

The following are the reliability measures used:

- PROCR: The probability that all operative repeaters can communicate with the station.

- EFNCR: The expected fraction of repeater pairs that can communicate through the station.

- PER(L): The probability that a repeater on level L (L hops away) can communicate with the station.

4.4.2 Effect of Routing Strategy

Our preconceived opinion of the process was that tree routing was certainly very unreliable since the failure of only one repeater could cut off a group of repeaters from the station. In addition, we felt that restricted routing would provide a great improvement over tree routing and nearly achieve the performance of the most reliable strategy, the adaptive routing. Restricted routing is preferred for reasons not having to do with reliability. Our results indicate that restricted routing reduces the gap between tree and adaptive routing global reliability by 50%. In this case, our measures of global reliability were the probability that all operating repeaters can communicate with the station, PROCR, and the expected fraction of repeater pairs communicating through the station, EFNCR. The reduction was in this range for both measures under all changes in the station node and at all failure probability levels. Figures 4.5 through 4.10 illustrate these results.

To measure the performance of repeaters on each level we computed average probability that a repeater on that hierarchical level could communicate with the station. For levels 2 or 3 hops away, restricted routing reduced the gap between tree and adaptive by less than 50%, however for levels further out its performance was better. For levels 5 or 6 hops away, it reduced the gap by well over 50%. In fact, it appears that for levels many hops away the performance of restricted and adaptive routing may be indistinguishable. Figures 4.11 through 4.13 illustrate these results.

As should be evident, these results have changed some of our earlier opinions. Restricted routing does not produce reliability values close to adaptive. It cuts in half the gap between tree and adaptive, however this improvement is disappointing considering how unreliable tree routing can be.

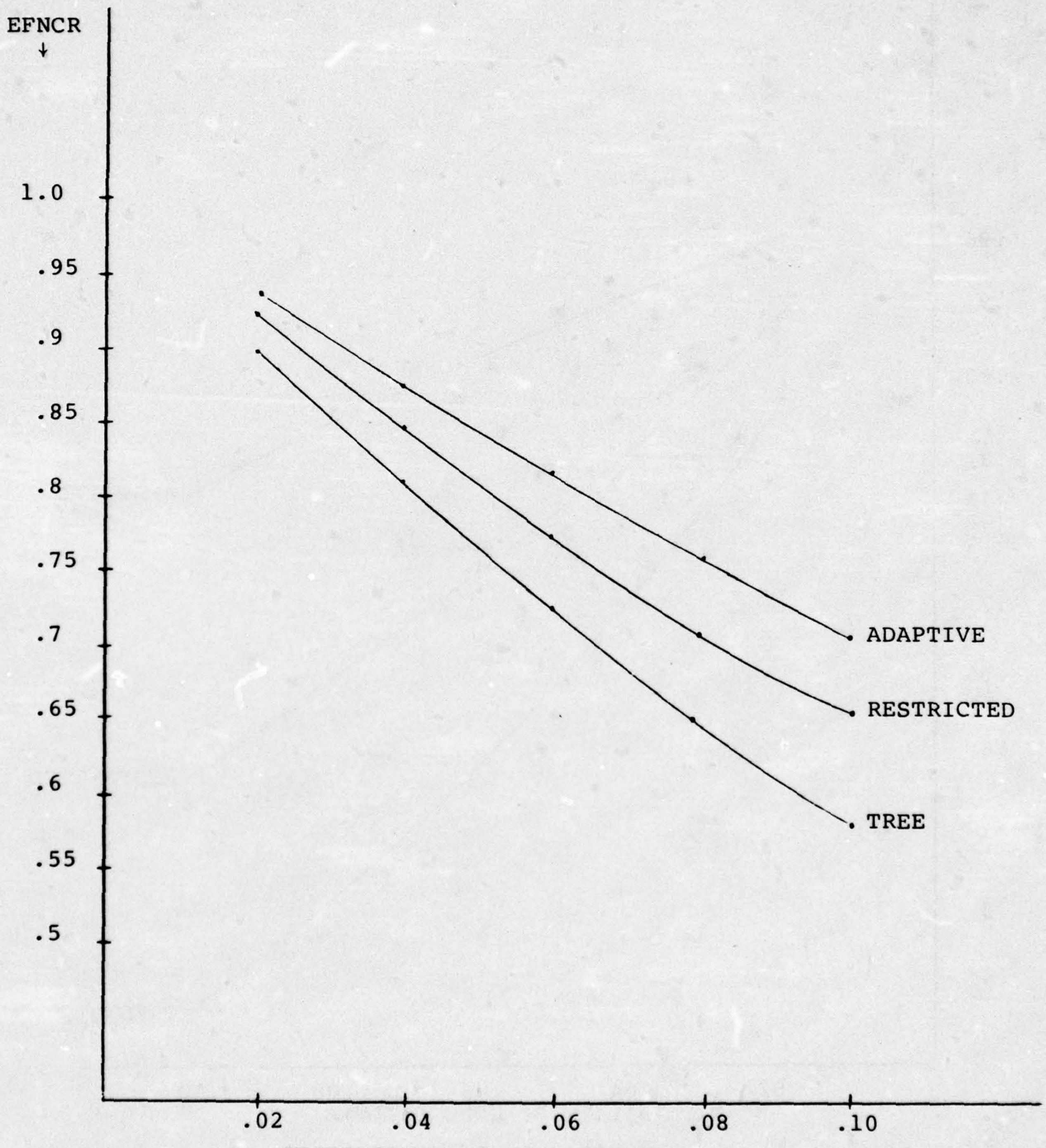


FIGURE 4.5: EXPECTED FRACTION OF NODE PAIRS THAT CAN COMMUNICATE THROUGH THE ROOT AS A FUNCTION OF NODE FAILURE, FOR DIFFERENT ROUTING ALGORITHMS AND FOR A NETWORK WITH A ROOT NODE IN CORNER

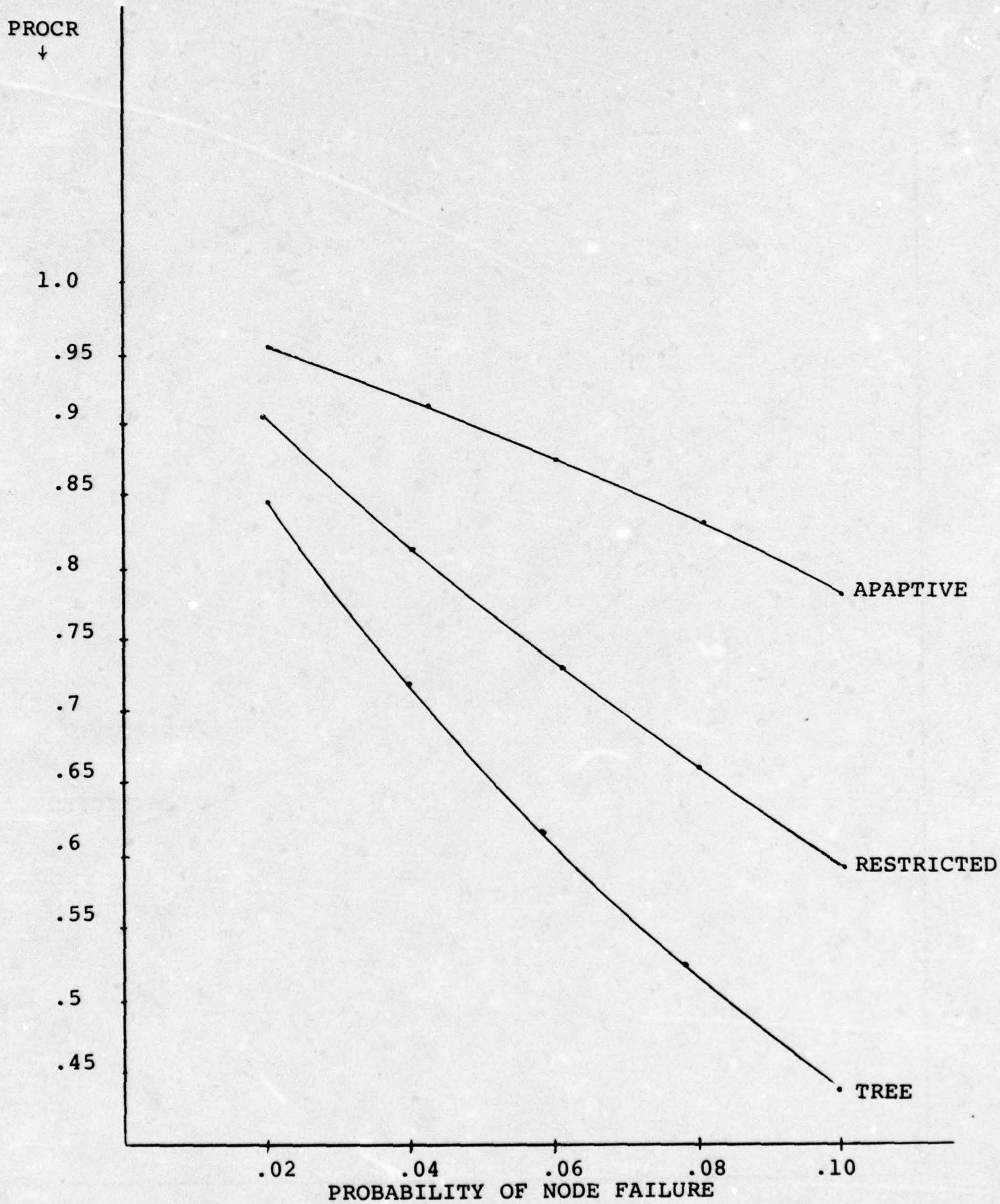


FIGURE 4.6: PROBABILITY THAT OPERATING NODES CAN COMMUNICATE WITH THE ROOT AS A FUNCTION OF NODE FAILURE, FOR DIFFERENT ROUTING ALGORITHMS AND FOR A NETWORK WITH A ROOT NODE IN CORNER

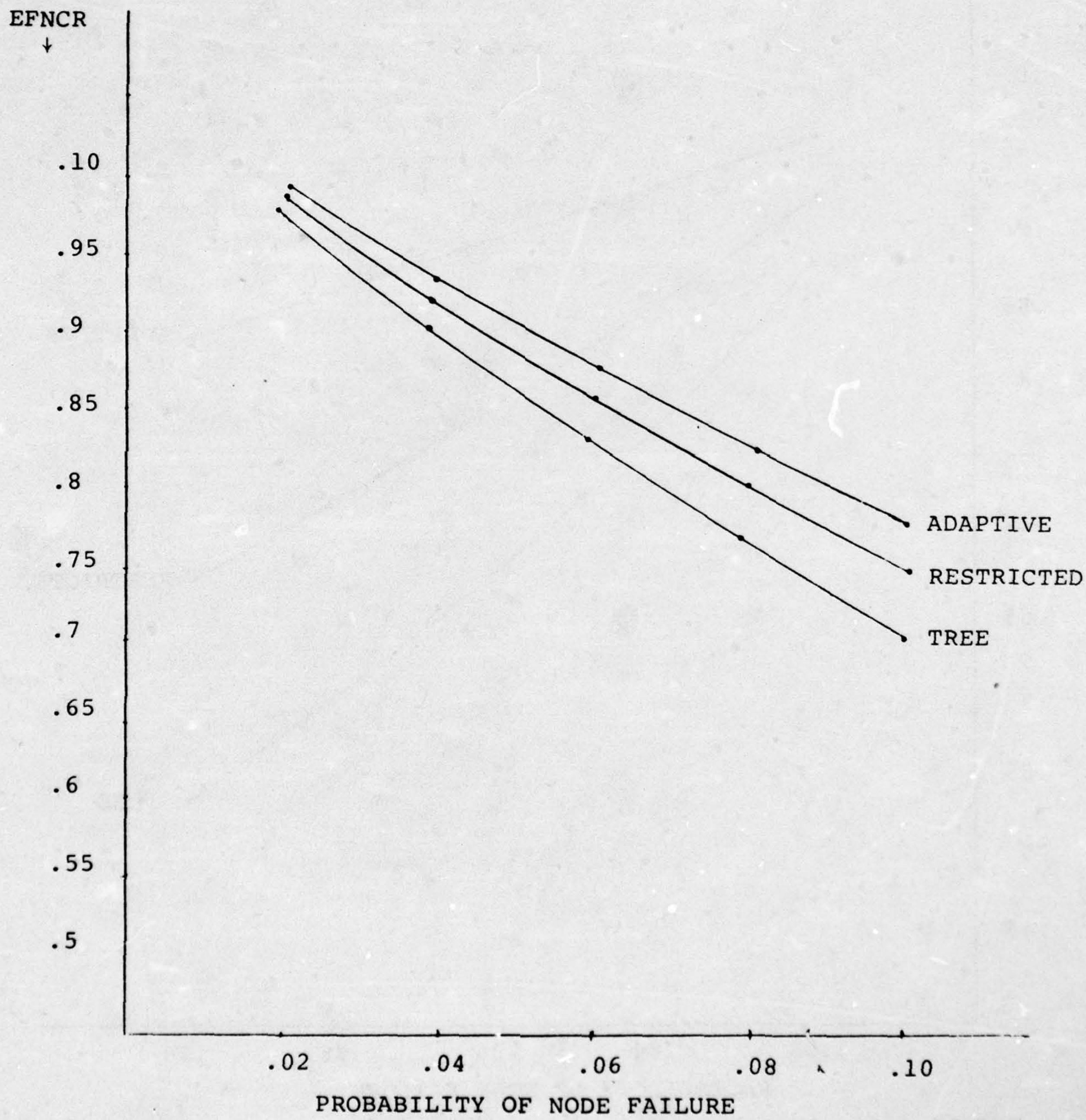


FIGURE 4.7: EXPECTED FRACTION OF NODE PAIRS THAT CAN COMMUNICATE THROUGH THE ROOT AS A FUNCTION OF NODE FAILURE, FOR DIFFERENT ROUTING ALGORITHMS AND FOR A NETWORK WITH TWO ROOT NODES

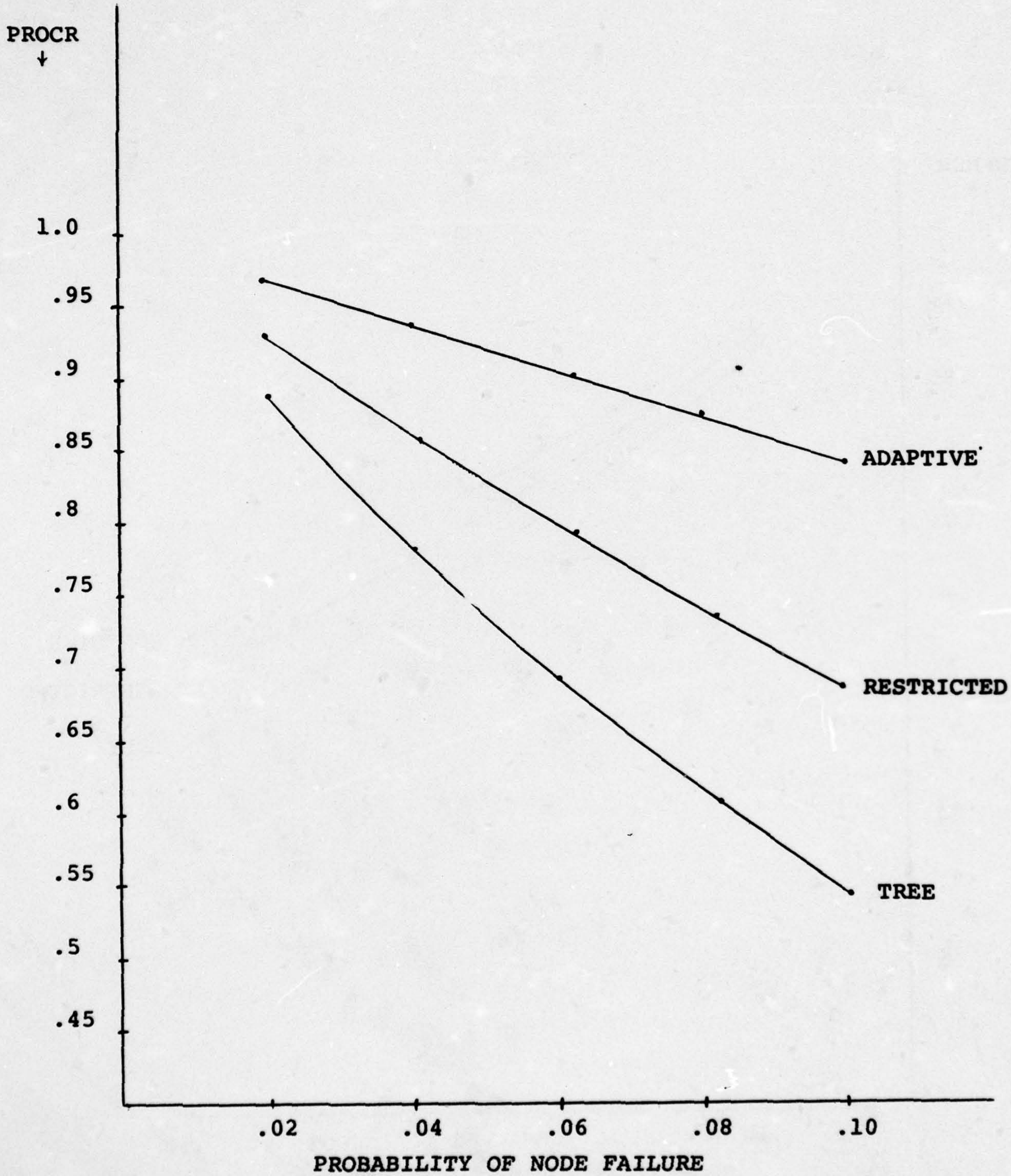


FIGURE 4.8: PROBABILITY THAT OPERATING NODES CAN COMMUNICATE WITH THE ROOT AS A FUNCTION OF NODE FAILURE, FOR DIFFERENT ROUTING ALGORITHMS AND FOR A NETWORK WITH TWO ROOT NODES

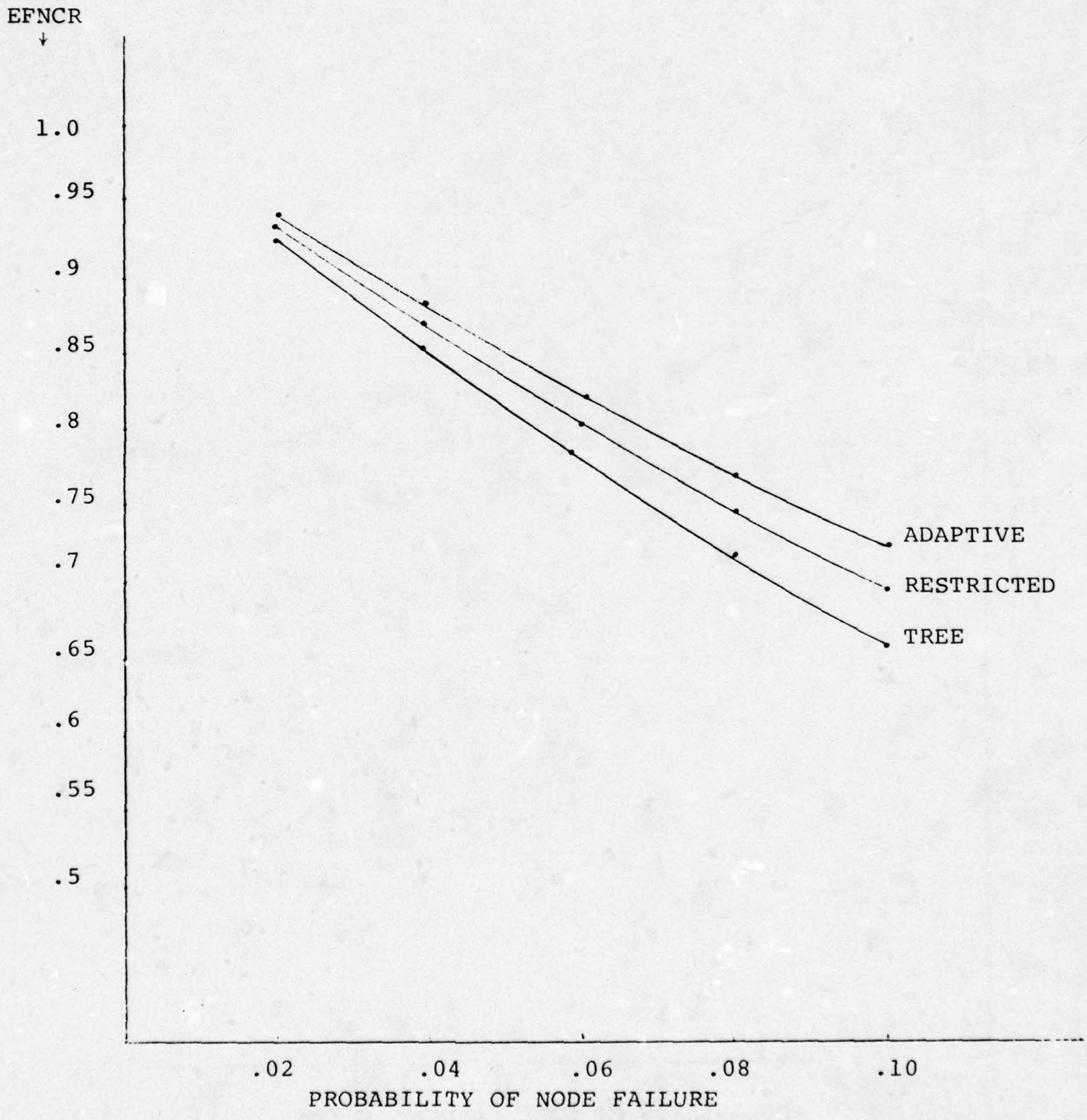


FIGURE 4.9: EXPECTED FRACTION OF NODE PAIRS THAT CAN COMMUNICATE THROUGH THE ROOT AS A FUNCTION OF NODE FAILURE, FOR DIFFERENT ROUTING ALGORITHMS AND FOR A NETWORK WITH ROOT NODE IN THE CENTER

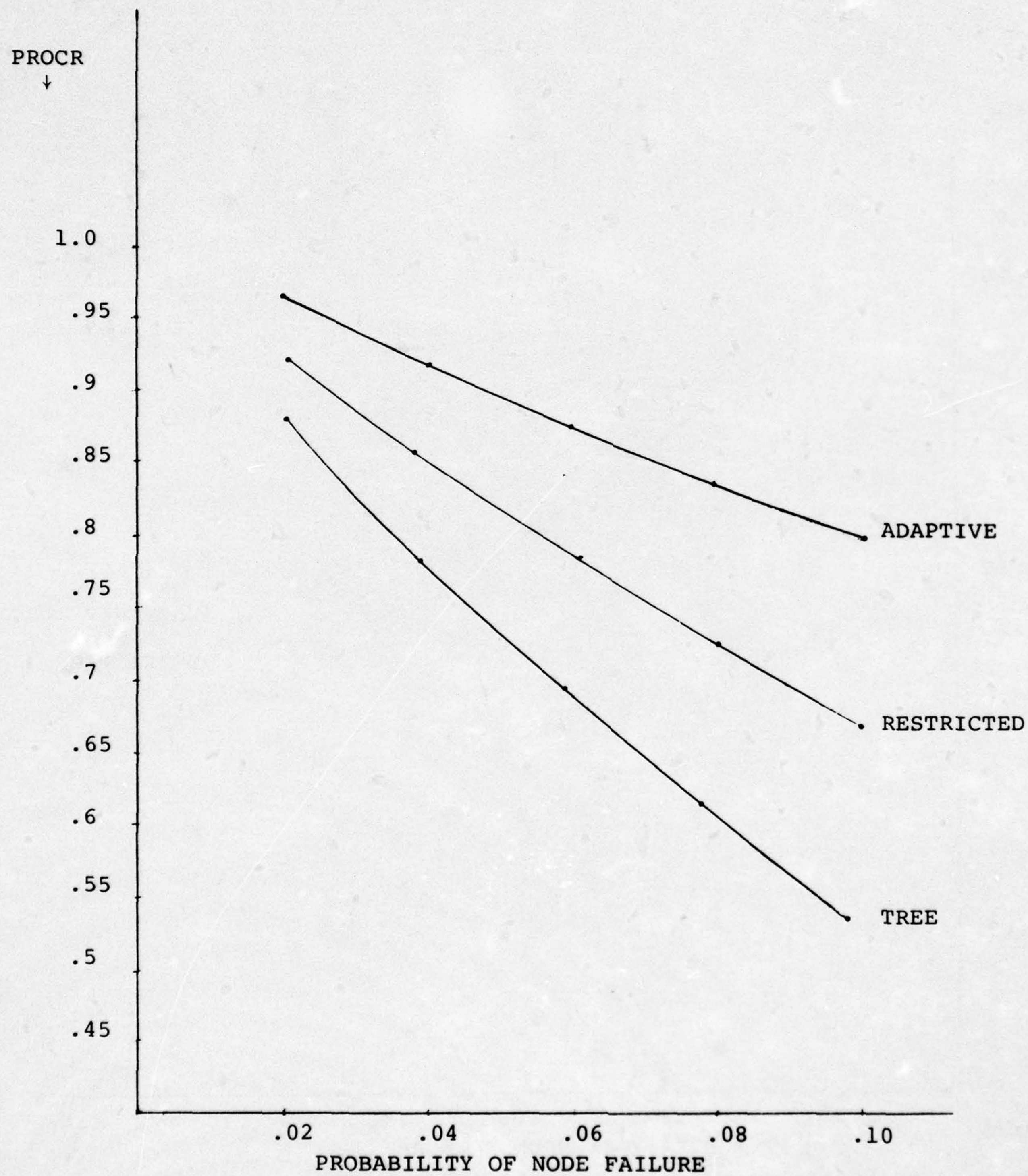


FIGURE 4.10: PROBABILITY THAT OPERATING NODES CAN COMMUNICATE WITH THE ROOT AS A FUNCTION OF NODE FAILURE, FOR DIFFERENT ROUTING ALGORITHMS AND FOR A NETWORK WITH ROOT NODE IN THE CENTER

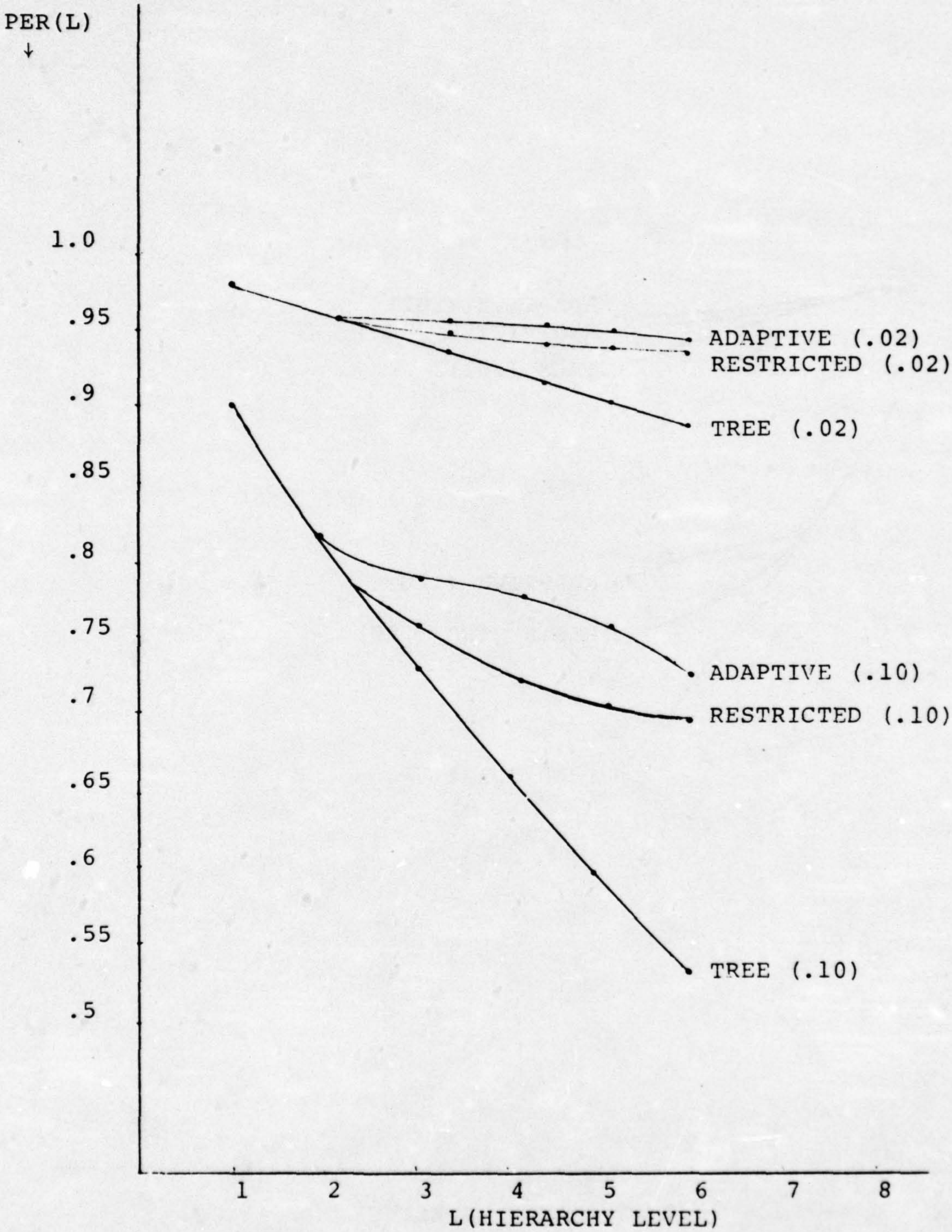


FIGURE 4.11: PROBABILITY THAT A NODE ON LEVEL L CAN COMMUNICATE WITH THE ROOT AS A FUNCTION OF L FOR DIFFERENT ROUTING ALGORITHMS AND TWO VALUES OF NODE FAILURE, FOR A NETWORK WITH THE ROOT NODE IN THE CORNER

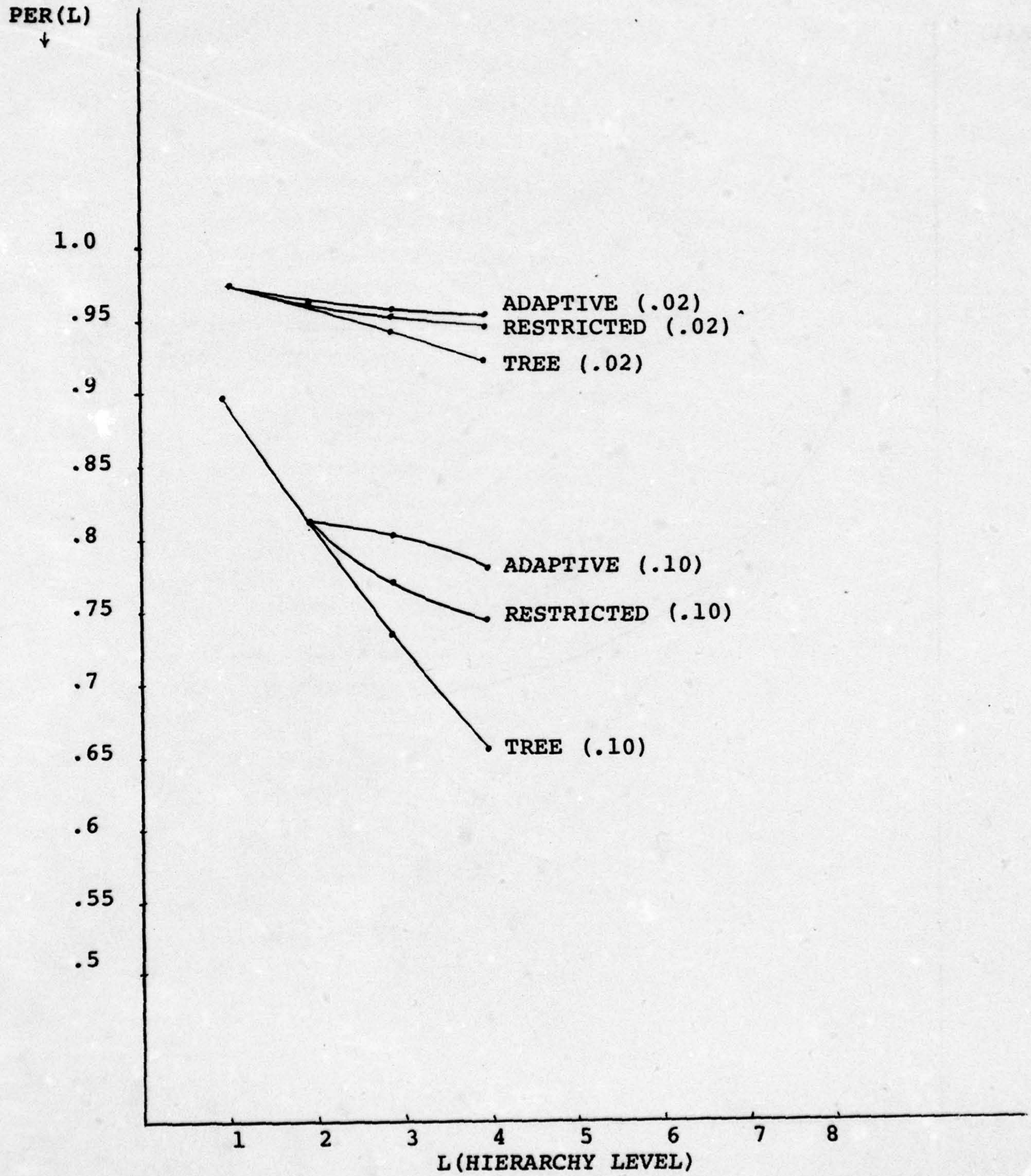


FIGURE 4.12: PROBABILITY THAT A NODE ON LEVEL L CAN COMMUNICATE WITH THE ROOT AS A FUNCTION OF L FOR DIFFERENT ROUTING ALGORITHMS AND TWO VALUES OF NODE FAILURE, FOR A NETWORK WITH TWO ROOT NODES

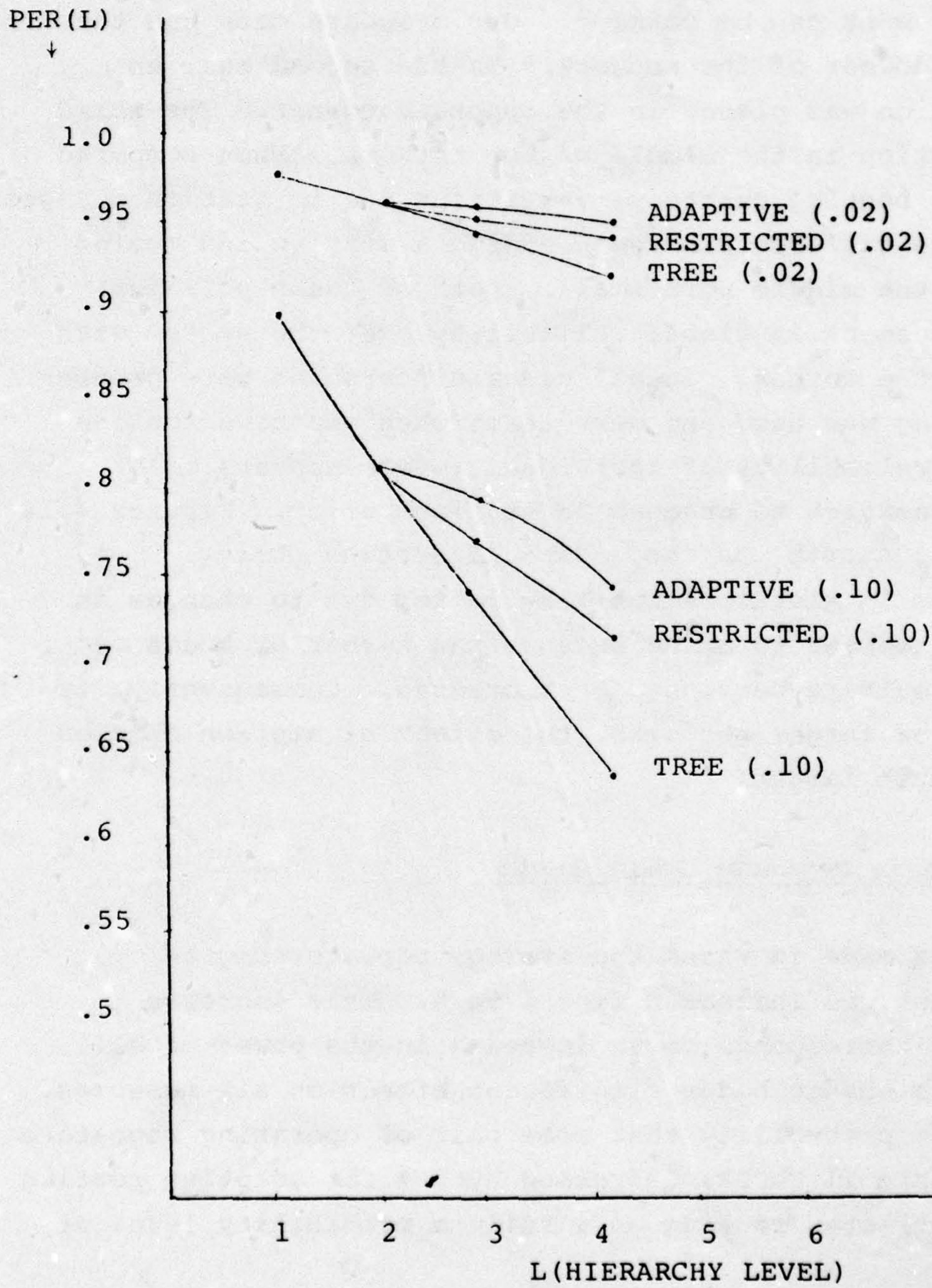


FIGURE 4.13: PROBABILITY THAT A NODE ON LEVEL L CAN COMMUNICATE WITH THE ROOT AS A FUNCTION OF L FOR DIFFERENT ROUTING ALGORITHMS AND TWO VALUES OF NODE FAILURE, FOR A NETWORK WIT[H] THE ROOT NODE IN THE CENTER

4.4.3 Variation In The Number of Stations and Station Placement

To increase reliability an additional station can be added or station placement can be changed. Our standard case had the station in the corner of the network. In the second run, an additional station was placed in the opposite corner. The third run had one station in the middle of the network. When compared to variation in routing strategy, variations due to station changes were small. The difference between adding a station and moving the station to the middle were small. Both of these policies showed an improvement in global reliability over the system with one station in the corner. In all cases differences were greater when tree routing was used and were least when adaptive routing was used. The reliability of individual levels appears to be relatively insensitive to changes in station policy. Figures 4.14 through 4.18 illustrate the comparison in station policy.

The changes in global reliability values due to changes in station policy, appear to arise because the number of nodes on levels further out are decreased or increased. Consequently, if we had considered larger networks, the effect of station changes may have been much larger.

4.4.4 Change in Repeater Power Level

One run was made in which the average repeater degree (number of links) was increased from 4 to 6. This increase in network density corresponds to an increase in the power of all repeaters. This change had a significant effect on all measures. For example, the probability that some pair of operating repeaters cannot communicate (1-PROCR) decreased by 50% for adaptive routing and 40% for restricted routing at a failure probability level of

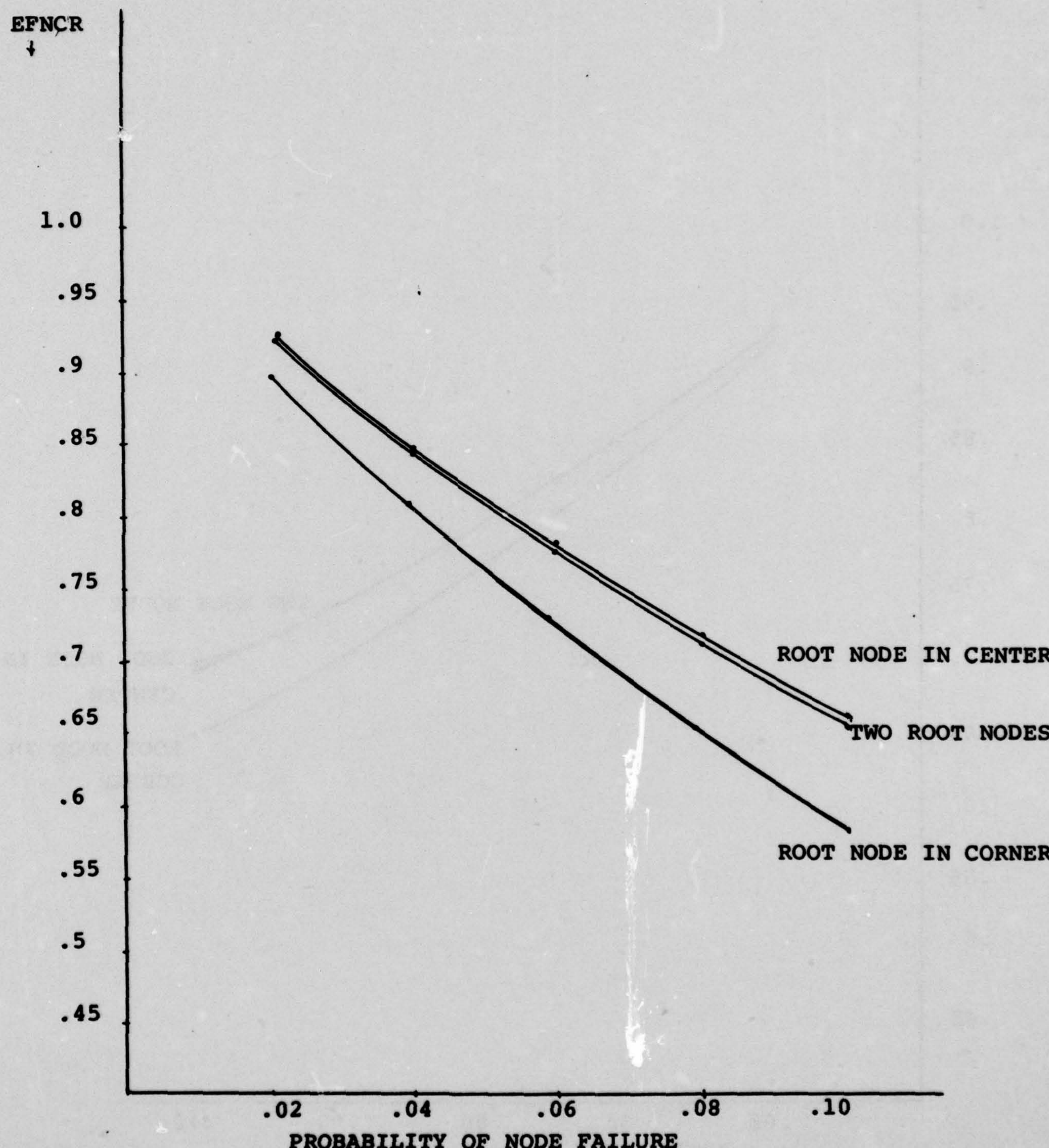


FIGURE 4.14: EXPECTED FRACTION OF NODE PAIRS THAT CAN COMMUNICATE THROUGH THE ROOT FOR TREE ROUTING AS A FUNCTION OF NODE FAILURE FOR DIFFERENT NETWORK TOPOLOGIES

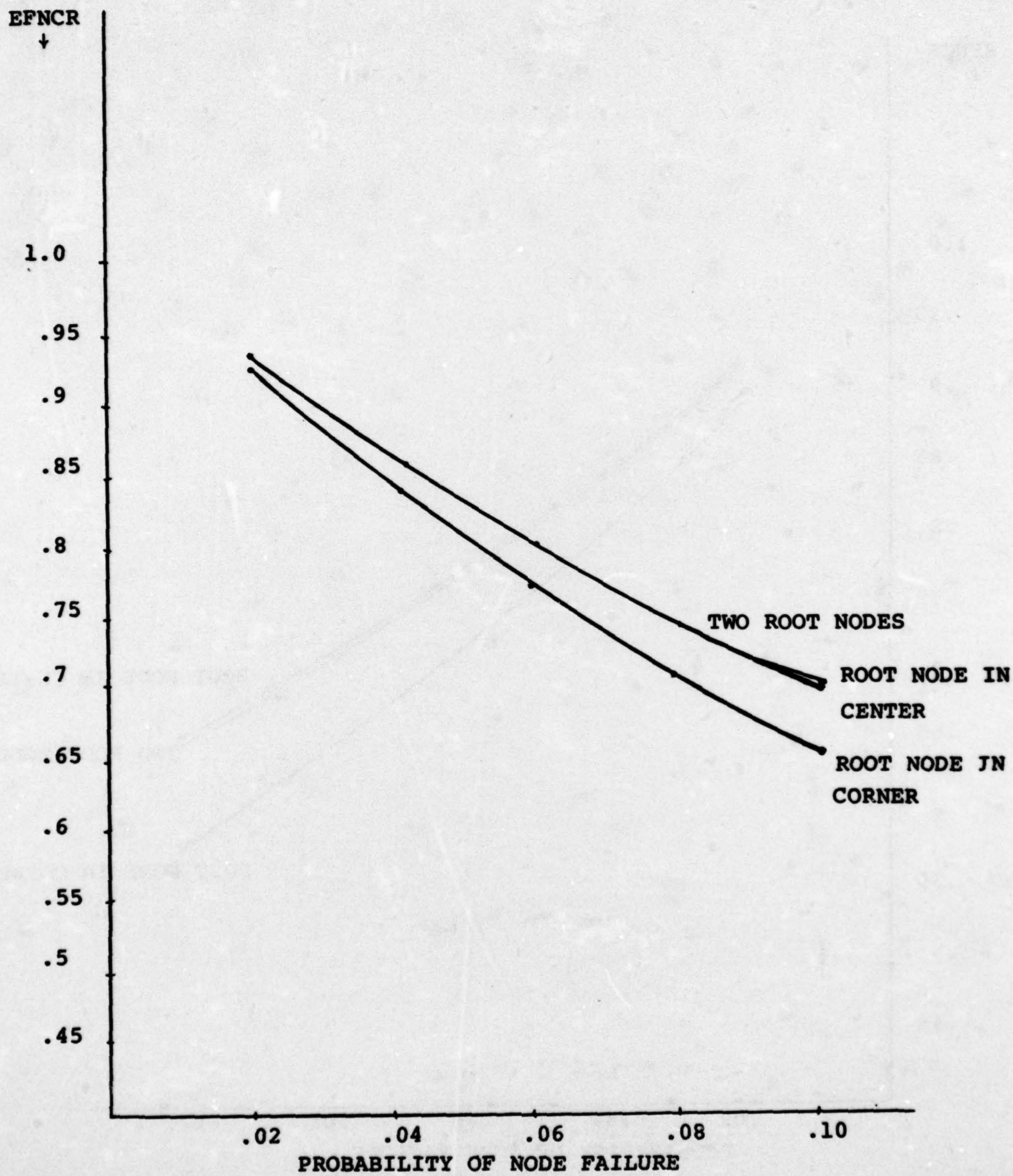


FIGURE 4.15: EXPECTED FRACTION OF NODE PAIRS THAT CAN COMMUNICATE THROUGH THE ROOT FOR RESTRICTED ROUTING AS A FUNCTION OF NODE FAILURE FOR DIFFERENT NETWORK TOPOLOGIES

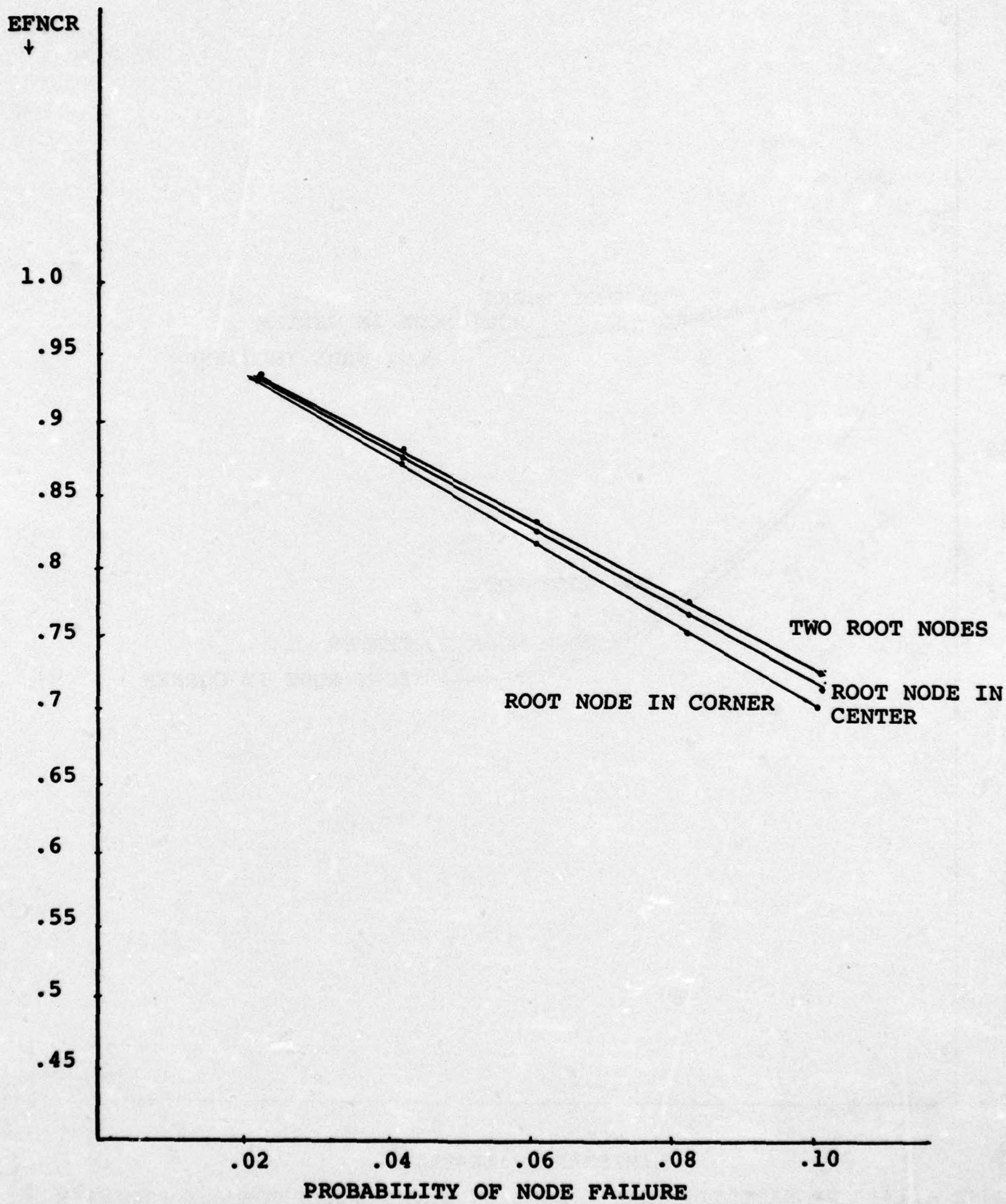


FIGURE 4.16: EXPECTED FRACTION OF NODE PAIRS THAT CAN COMMUNICATE THROUGH THE ROOT FOR ADAPTIVE ROUTING AS A FUNCTION OF NODE FAILURE FOR DIFFERENT NETWORK TOPOLOGIES

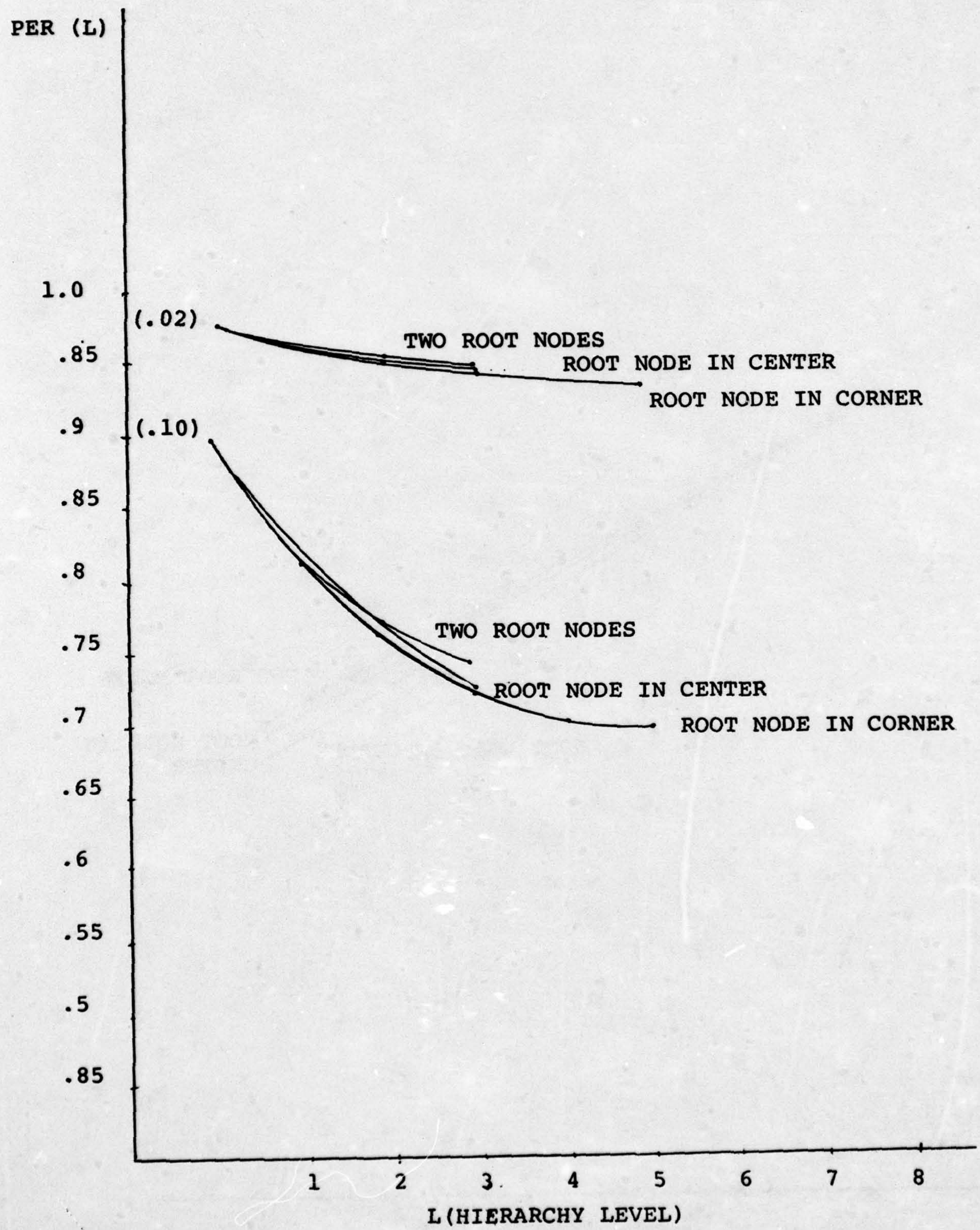


FIGURE 4.17: PROBABILITY THAT A NODE ON LEVEL L CAN COMMUNICATE WITH THE ROOT FOR RESTRICTED ROUTING AS A FUNCTION OF L FOR TWO VALUES OF FAILURE AND FOR DIFFERENT NETWORK TOPOLOGIES

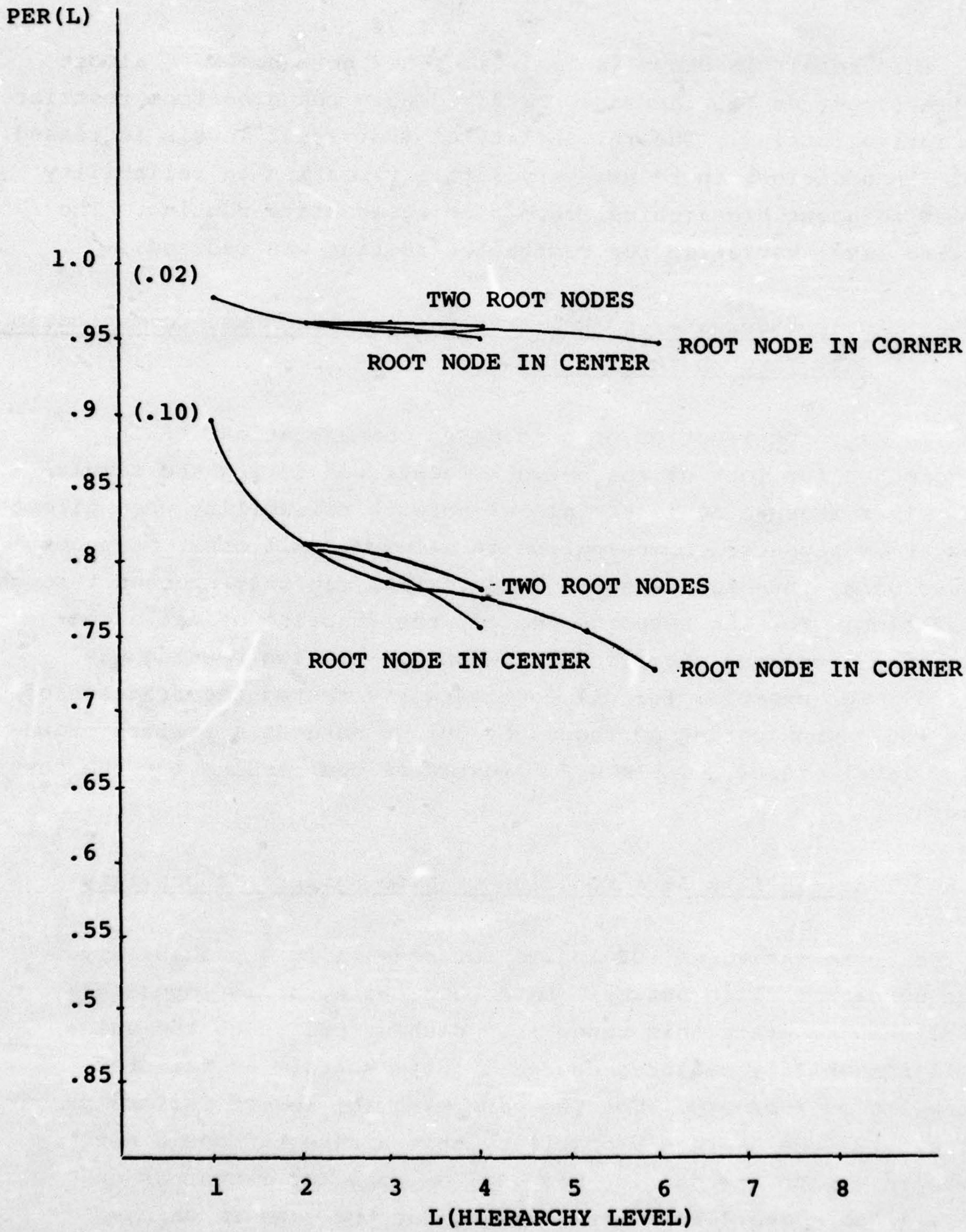


FIGURE 4.18: PROBABILITY THAT A NODE ON LEVEL L CAN COMMUNICATE WITH THE ROOT FOR ADAPTIVE ROUTING AS A FUNCTION OF L FOR TWO VALUES OF FAILURE AND FOR DIFFERENT NETWORK TOPOLOGIES

.02. This result is shown in Table 1. The increases were almost as significant as the increases obtained when changing from restricted to adaptive routing. The reliability of individual levels increased also. In addition, there was very little variation in reliability between adjacent hierarchical levels using adaptive routing. The level to level variation for restricted routing was reduced.

4.4.5 Direct Repeater-to-Repeater Communication Vs. Communication Only Through the Station

The expected fraction of node pairs communicating, EFNCR, was computed for four of the networks generated during the simulation. This measure evaluates global network reliability when direct repeater-to-repeater communication is allowed. All other measures we have used, thus far, assume communication can only proceed through the station. For the networks tested, the fraction of pairs communicating increased significantly when non-station routing was allowed. For example, for all four networks tested the fraction of pairs not communicating decreased by 30% or more at a failure probability level of .02. Table 4.2 illustrates the results for the test networks.

4.4.6 Reliability As a Function of Node Failure Probability

We chose the range .02 to .10 for node failure probability. Since no packet radio networks have been built, it is impossible to tell how accurate this range is. Within this range the value of all reliability measures decrease quite sharply as failure probabilities increase. For the case with the lowest failure probability, .02, the average probability that a repeater could not communicate with the station was .045 or more for repeaters on levels 4 and above. This was obtained for the case in which

TREE		Ave. Degree=4		Ave. Degree=6	
	PROB	PROCR	EFNCR	PROCR	EFNCR
	.02	.848	.899	.904	.921
	.04	.718	.808	.817	.847
	.06	.608	.724	.756	.779
	.08	.514	.649	.687	.715
	.10	.435	.580	.624	.656

RESTRICTED		Ave. Degree=4		Ave. Degree=6	
	PROB	PROCR	EFNCR	PROCR	EFNCR
	.02	.903	.922	.942	.935
	.04	.815	.848	.886	.872
	.06	.735	.779	.833	.812
	.08	.662	.714	.783	.755
	.10	.596	.653	.736	.701

ADAPTIVE		Ave. Degree=4		Ave. Degree=6	
	PROB	PROCR	EFNCR	PROCR	EFNCR
	.02	.960	.937	.980	.944
	.04	.918	.876	.960	.890
	.06	.876	.817	.940	.838
	.08	.833	.760	.919	.787
	.10	.789	.705	.898	.739

TABLE 4.1: RELIABILITY VALUES FOR DIFFERENT TRANSMISSION POWER LEVELS

	NET 1		NET 2	
PROB	EFNCR	EFNC	EFNCR	EFNC
.02	.943	.960	.918	.946
.04	.887	.919	.840	.894
.06	.832	.878	.767	.843
.08	.779	.837	.698	.794
.10	.728	.796	.633	.745

	NET 3		NET 4	
PROB	EFNCR	EFNC	EFNCR	EFNC
.02	.944	.960	.943	.960
.04	.889	.921	.888	.920
.06	.836	.883	.834	.880
.08	.785	.844	.782	.841
.10	.735	.806	.731	.802

TABLE 4.2: ADAPTIVE ROUTING*

* Net 2 was sparser than the other nets.

the station is placed in the corner. This level of reliability would not seem to be acceptable. Consequently, failure probability less than .02 may be necessary. Of course, reliability levels can be increased by other mechanisms such as those mentioned earlier in this section.

4.4.7 Variations in Node Failure Probability

In all of our studies, we made the failure probabilities of all repeaters equal. In reality failure probabilities may vary from repeater to repeater. To evaluate the effect of such variations we made four runs in which repeater and station failure probabilities were randomly chosen. The probabilities were chosen so that the average equaled the values we used in previous runs. The values of global reliability measures varied by at most 3% from those in previous runs, and in the majority of the cases, the variation was less than 1%. While these percentages may seem insignificant, they could be bothersome since in many cases the important value is one minus the reliability measured. For this value, the percentages will be higher. However, we have noticed that the most significant part of the variation comes from variations in the station failure probability. Presumably, this value would be estimated independently from repeater value. Consequently, we feel that results of analysis using constant repeater failure probabilities and possibly a different station failure probability adequately approximate the real situation.

4.5 GENERAL CONCLUSION

We found packet radio network reliability to be surprisingly sensitive to variation in a number of design parameters. Using adaptive routing rather than restrictive, increasing repeater power levels, allowing direct repeater-to-repeater communication, and decreasing repeater failure probabilities can all drastically increase the reliability. In addition, trends from this study indicate that for larger networks the number and locations of the stations can also effect reliability. It also appears that even with repeater failure probabilities as low as .02 reliability levels may not be high enough in our standard case (when the station is in the corner). Consequently, some of the options mentioned above will have to be implemented in order to increase reliability.

The strongest conclusion of this study is that reliability must be explicitly considered when designing packet radio networks. An effective design procedure should look at a large range of network and routing parameters. The reliability associated with each combination of parameters should be evaluated. The final design can then be chosen from among those designs with acceptable reliability levels.

VOLUME 2

Chapter 4

REFERENCES

REFERENCES

[BALL, 1975] Ball, M., and R. M. Van Slyke, "Backtracking Algorithms for Network Reliability Analysis," Technical Report No. 015-01, Network Analysis Corporation, August, 1975.

[FRANK, 1975] Frank, H., I. Gitman, and R. M. Van Slyke, "Packet Radio System - Network Considerations," Proceedings of the National Computer Conference, May, 1975, pp. 217-231.

[GITMAN, 1976] Gitman, I., R. M. Van Slyke, and H. Frank, "Routing in Packet Switching Broadcast Radio Networks," IEEE Trans. on Communications, August 1976.

[KAHN, 1975] Kahn, R. E., "The Organization of Computer Resources into a Packet Radio Network," Proceedings of the National Computer Conference, May, 1975, pp. 177-186.

[NAC, 1974] (Network Analysis Corporation), The Practical Impact of Recent Computer Advances on the Analysis and Design of Large Scale Networks, Third Semiannual Technical Report, June, 1974, available from Defense Documentation Center, Arlington, Virginia.

[NAC, 1976] (Network Analysis Corporation), "Computational Techniques," Sixth Semiannual Technical Report, Vol. 5, February, 1976, available from Defense Documentation Center, Arlington, Virginia.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) NETWORK ANALYSIS CORPORATION BEECHWOOD, OLD TAPPAN ROAD GLEN COVE, NEW YORK 11542	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP None

3. REPORT TITLE
SIXTH SEMIANNUAL TECHNICAL REPORT, FEBRUARY 1976, FOR THE PROJECT, "LOCAL, REGIONAL AND LARGE SCALE INTEGRATED NETWORKS," (VOLUMES 1 THROUGH 5)

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
SIXTH SEMINANNUAL REPORT, FEBRUARY 1976

5. AUTHOR(S) (Last name, first name, initial)
NETWORK ANALYSIS CORPORATION
BEECHWOOD, OLD TAPPAN ROAD
GLEN COVE, NEW YORK 11542

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
FEBRUARY 1976	880	140

8a. CONTRACT OR GRANT NO. DAHC-15-73-C0135 b. PROJECT NO. c. ARPA ORDER NO. 2286 d.	9a. ORIGINATOR'S REPORT NUMBER(S) SEMIANNUAL REPORT 6 (5 VOLUMES) 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) ---

10. AVAILABILITY/LIMITATION NOTICES

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
None	Advanced Research Projects Agency, Department of Defense

13. ABSTRACT This technical report presents significant accomplishments in the development of computational techniques for network analysis and design, and the following cost/performance tradeoffs for data and computer communication networks: cost-effectiveness and economies of scale previously demonstrated for medium size packet switching networks (e.g., ARPANET) extend to large scale networks (e.g., 10 Mbs throughput). The most cost-effective architecture for large scale networks is a multi-level hierarchical structure. The cost of hierarchical networks is not very sensitive to number of backbone nodes near the optimum; the local access cost is approximately 50% of the total, and is the most sensitive component. For AUTODIN II, Phase I, the minimum cost backbone network configuration is a hybrid terrestrial-satellite net with six (out of eight nodes) ground stations, resulting in 25% savings in backbone line cost compared to a terrestrial net only. Packet Radio local access is a cost-effective alternative to conventional techniques, providing a Packet Radio Unit can be manufactured for $20K-$30K. Significant results are reported on Packet Radio network communication protocols, throughput-delay-reliability analysis, and stability control.

14. KEY WORDS Computer networks, communication networks, terrestrial and satellite networks, packet radio networks, throughput, cost, delay, blocking, ARPA Computer Network, store-and-forward, packet switching.

ABSTRACT

UNCLASSIFIED
Security Classification